A Publication of the American Psychological Association

The Journal of Abnormal and Social Psychology

Founded by Morton Prince, 1906

CONTENTS

ARTICLES

Interrelations Among Three Criteria of Improvement in Group Therapy: Comfort, Effectiveness, and Self-awareness 281
Herbert C. Kelman and Morris B. Parloff

Determinants of Conformity and Attitude Change 289
Kenneth R. Hardy

Test-Retest Stability of Interaction Patterns During Interviews Conducted One Week Apart 295
George Saslow, Joseph D. Matarazzo, Jeanne S. Phillips, and Ruth G. Matarazzo

Authoritarian Attitudes and Personality Maladjustment 303
Arthur R. Jensen

Factors Determining the Guessing of Meanings of Foreign Words 312
Yvonne Brackbill and Kenneth B. Little

Performance on the Bender Under Hypnotic Age Regression 319
Harold B. Crasilneck and Carmen Miller Michael

Decision Processes in Communication Nets 323
Marvin E. Shaw, Gerard H. Rothschild, and John F. Strickland

Continued on inside front cover

This is the Last Issue of Volume 54.
Title Page and Contents for the Volume
Appear Herein.

Volume 54 MAY 1957 Number 3

CONTENTS—*Continued*

A Cross-Cultural Study of Perceptual Predominance in Binocular Rivalry . . . 331
James W. Bagby
Psychogenic Deafness and Perceptual Defense . . . 335
Milton L. Klein man
Expectancy Changes in Skill and Chance Situations . . . 339
E. Jerry Phares
Sex Differences in Social Memory Tasks . . . 343
Sam L. Witryol and Walter A. Kaess
Metaphors from Music Criticism . . . 347
Roger W. Brown, Raymond A. Leiter, and Donald C. Hildum
Liking for the Group and the Perceived Merit of the Group's Behavior . . . 353
Leonard Berkowitz
Studies in the Phenomenology of Emotions . . . 358
Jack Block
Effects of Biological and Social Motivation on Schizophrenic Reaction Time . 364
Gerald Rosenbaum, William R. Markaney, and James L. Grisell
Conclusion-Drawing, Communicator Credibility, and Anxiety as Factors in Opinion Change . . . 369
Bernard J. Fine
The Role of Experience in Judging Children's Photographs . . . 375
William Kessen
Relationships Between Child Training and the Pictorial Arts . . . 380
Herbert Barry, III
The Generality of Accuracy and Response Sets in Interpersonal Perception . . . 384
Wayman J. Crow and Kenneth R. Hammond
Patterns of Attitudes Toward Child Rearing and the Family . . . 391
Earl S. Schaefer and Richard Q. Bell
Authoritarian Ideology in Negro College Students . . . 396
George A. Stechler
Personality, Anxiety, and Muscle Tension as Determinants of Response Stereotypy . . . 400
James L. Kuethe and Charles W. Erikson
Leading and Following Psychotherapeutic Techniques with Hypnotically Induced Repression and Hostility . . . 405
Jesse E. Gordon
The Effects of Personal and Shared Threats Upon Social Prejudice . . . 411
Seymour Feshbach and Robert Singer

CASE REPORT

The Wolf . . . 417
R. K. Greenbank

CRITIQUE AND NOTES

A Guttman Scale for Sexual Experience—A Methodological Note . . . 420
Lawrence Podell and John C. Perkins
Levels of Perceptual Organization and Cognition: Conflicting Evidence . . . 422
E. W. Pickrel
Authoritarianism or Acquiescence in Bass's Data . . . 424
Samuel Messick and Douglas N. Jackson
Reply to Messick and Jackson's Comments on Authoritarianism or Acquiescence . . . 426
Bernard M. Bass

ERRATA . . . 428

Volume 54 MAY, 1957 Number 3

The Journal of Abnormal and Social Psychology

Editor: M. BREWSTER SMITH, *New York University*

Associate Editor: E. J. SHOBEN, JR., *Teachers College, Columbia University*

CONSULTING EDITORS

GORDON W. ALLPORT
Harvard University

URIE BRONFENBRENNER
Cornell University

DONALD T. CAMPBELL
Northwestern University

LAUNOR F. CARTER
The RAND Corporation

RICHARD CHRISTIE
Columbia University

MORTON DEUTSCH
Bell Telephone Laboratories

I. E. FARBER
State University of Iowa

LEON FESTINGER
Stanford University

DONALD W. FISKE
University of Chicago

ROBERT R. HOLT
New York University

HERBERT H. HYMAN
Columbia University

IRVING L. JANIS
Yale University

HAROLD E. JONES
University of California

HAROLD H. KELLEY
Center for Advanced Study in the Behavioral Sciences

LEONARD S. KOGAN
Community Service Society of New York

GARDNER LINDZEY
Syracuse University

DAVID C. MCCLELLAND
Harvard University

QUINN MCNEMAR
Stanford University

LEO POSTMAN
University of California

NEVITT SANFORD
University of California

SEYMOUR SARASON
Yale University

RICHARD L. SOLOMON
Harvard University

RENATO TAGIURI
Harvard University

Managing Editor: ARTHUR C. HOFFMAN

Assistant Managing Editor: HELEN ORR

Editorial Staff: BARBARA CUMMINGS, FRANCES H. CLARK, SADIE J. DOYLE

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis, those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued bimonthly in January, March, May, July, September, and November, two volumes per year. The two volumes average 784 pages. The subscription price per year is $16.00; foreign, $16.50. Single issues are $3.00.

Original manuscripts should be sent to the Editor:

DR. M. BREWSTER SMITH
Graduate Department of Psychology
New York University
Washington Square
New York 3, N. Y.

Fifty free offprints are supplied to contributors. Authors of prior publication articles receive no gratis offprints.

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street N.W.
Washington 6, D. C.

Changes of address must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Printed and mailed by:

WAVERLY PRESS, INC.
Baltimore 2, Maryland

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Mt. Royal and Guilford Aves., Baltimore 2, Md.
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Entered as second-class matter February 15, 1954 at the Post Office at Baltimore, Maryland under the Act of February 28, 1925. Acceptance for mailing at special rate of postage provided for in Section 34.40 (e), P. L. and R. Additional entry at Washington, D.C.

Copyright 1957 by the American Psychological Association, Inc.

INTERRELATIONS AMONG THREE CRITERIA OF IMPROVEMENT IN GROUP THERAPY: COMFORT, EFFECTIVENESS, AND SELF-AWARENESS[1]

HERBERT C. KELMAN AND MORRIS B. PARLOFF

National Institute of Mental Health, U. S. Department of Health, Education and Welfare

THE widespread use of global evaluations of improvement in psychotherapy assumes that improvement is a unitary phenomenon, i.e., that the different criteria of improvement vary together. Clinical evidence suggests that this assumption may be unwarranted. It becomes necessary, therefore, to spell out separate criteria of improvement and to develop measures for each. Only when specific criteria and objective measures are available will it be possible to gain some understanding of the nature of changes produced in therapy and the process by which such changes are effected.

Criteria of improvement have been proposed by Kogan and Hunt (6), who distinguish three points of view from which psychotherapy is evaluated: that of the person who receives the psychotherapy, that of the other persons who come in contact with him, and that of mankind more generally. They suggest another way of distinguishing criteria of improvement when they write that, "Pervading all of the theories [of psychotherapy] either explicitly or implicitly is the conception that the primary aim of therapy is improved adjustment of the individual, his level of adjustment being a function of his understanding of the self and the world, his handling of anxiety, and the effectiveness and social acceptability of his behavior" (6, p. 132). Miller (7) lists four criteria that have been used in the evaluation of therapy: (*a*) the disappearance of symptoms, (*b*) the patient's social adjustment, (*c*) the patient's personal satisfaction, and (*d*) the therapist's satisfaction with the patient's adjustment. Miller considers these criteria as conflicting with one another, and therefore stresses the importance of studying to what extent and under what circumstances they are correlated.

Using the distinctions made by Kogan and Hunt and by Miller as a point of departure, the authors derived three criteria of improvement (10): 1. *Comfort:* The patient's stated purpose in coming to treatment is usually to increase his comfort in his daily functioning, and particularly to alleviate his symptoms. This conscious goal, of course, does not represent the complete picture of his motivations regarding therapy. Secondary gains derived from the illness, for example, may produce forces in the opposite direction. 2. *Effectiveness:* Those with whom the patient interacts in his daily life are usually most concerned with his effectiveness as a social being, his success in the performance of his social roles and in the management of his transactions with others. 3. *Self-awareness:* Most therapists, it can be assumed, are also concerned with increasing the patient's comfort and effectiveness. They tend to judge the effects of therapy, however, by reference to certain additional values, which are considered prerequisites to the attainment of greater comfort and effectiveness. These values include, for example, the acquisition of "insight," the correction of perceptual distortions, and the release of repressions. Self-awareness is one such value which is shared by a number of therapeutic approaches and was therefore selected as a third criterion of improvement.

It is reasonable to expect a certain amount of overlap between these three criteria, particularly since they may often involve evaluations of the same behavior from three different viewpoints. Nevertheless, it is quite likely that therapy will have differential effects on such

[1] The research for this paper was done at the Henry Phipps Psychiatric Clinic, Johns Hopkins Hospital as part of the Group Psychotherapy Research Project. Funds necessary to carry out the experimental work reported here were supplied by the Veterans Administration. The completion of the investigation was supported by a research grant from the National Institute of Mental Health. Herbert C. Kelman was a Public Health Service Research Fellow of the National Institute of Mental Health while participating in the study.

The writers are grateful to Jerome D. Frank, Director of the Group Psychotherapy Research Project, who made immeasurable contributions to this study and to the thinking behind it, and to Eduard Ascher, Stanley Imber, Earl Nash, David Rosenthal, and Anthony Stone, for their cooperation and suggestions in all stages of the research.

criteria. It cannot be simply assumed that a patient who improves in one respect necessarily improves in all other respects as well. Instead, it is necessary to establish the specific relationships between different kinds of changes and to study the conditions affecting these relationships. The criteria proposed here represent one way of classifying the different kinds of change that can be expected. The purpose of the present study was to investigate the interrelations among these criteria, using a small sample of patients participating in group therapy.

METHOD

Since the three criteria are very broad concepts, it is possible to devise a variety of measures for each, designed to tap different aspects of the criterion. It is conceivable that the kinds of relationships found differ with the specific operations defining any given criterion. In order, therefore, to represent some of the manifold aspects of each criterion, three measures of comfort, two measures of effectiveness, and two measures of self-awareness were used in the present study. It would be more accurate, consequently, to say that the purpose of the study was to investigate the interrelations among seven measures of improvement, representing (though not exhaustively) three separate criteria of improvement.

Subjects

The sample used in this study consisted of fifteen neurotic patients, seven men and eight women, who participated in one of three therapy groups at the Out-Patient department of the Henry Phipps Psychiatric Clinic for a period of at least 20 weeks. The therapeutic method was analytically oriented group therapy. Two of the groups were conducted by one psychiatrist and the third by another. Both therapists had had a minimum of 12 years of experience in therapy. Their orientation was eclectic, based largely on the theories of Sullivan, Freud, and Meyer. Each group met once a week for one hour. Further details about the sample, the form of therapy, and the therapists can be found elsewhere (9).

The patients were tested at the beginning of therapy and again after 20 weeks of treatment. The difference between the "before" score and the "after" score on each measure constitutes the index of change or improvement. A description of the seven measures follows.

Measures of Comfort

Self-satisfaction Q sort. Rogers (11) and others (e.g., 4) have used the degree of congruence between the patient's perceived self and ideal self, obtained by correlating two Q sorts, as an index of self-satisfaction. The same general method was used in this study, but special Q sort items, relevant to the group situation, were devised.[2] Using Bion's concepts (1) as modified by Rosenthal (12) and Stock (13), six drives were distinguished which underlie interaction in group situations: pairing, counter-pairing, fight, flight, dependence, and counter-dependence. For each drive, ten items were selected describing behavior of a patient in a group judged to be characteristic of this particular drive (e.g., Fight: "I prolong or intensify arguments." Counter-dependence: "I resist suggestions of the doctor"). The patients sorted the 60 items twice: once to describe their characteristic behavior in the therapy group, and once to describe how they would ideally wish to behave. For each Q sort, sums of scores for each drive were obtained and rank ordered.[3] The rank-order correlation between the Perceived Self Q sort and the Ideal Self Q sort of a given patient provided the measure of his satisfaction with his behavior in the group. The difference between the rho obtained after four weeks of therapy and the rho obtained after 20 weeks constitutes the measure of change in the patient's degree of self-satisfaction.

Symptom-disability checklist. A modification of the Cornell Index (16) was administered before therapy and again after 20 weeks. The list contained 41 physiological, behavioral, and affective symptoms. Patients were asked to check those symptoms which they experienced in the preceding month, and to rate on a three-point scale the degree to which each of these symptoms was disabling, i.e., interfered with their work, social activities, and over-all feeling of well-being. The total score was based on the number of symptoms checked, each weighted by its degree of disability. The difference between the "before" and "after" scores constitutes the measure of symptomatic relief. A patient does not improve in this measure simply by losing his presenting complaints, but by demonstrating a *net* decline in number and intensity of all symptoms.

Discomfort evaluation scale. This scale consisted of ten items, describing areas of the patient's behavior and especially of his interpersonal relationships in which he might experience discomfort, e.g., feelings of dependence, guilt, social uneasiness, etc. For each patient, these items were rated by each of three staff members who had seen him before therapy and again after 20 weeks: a psychiatrist conducting a psychiatric interview which concentrated on the patient's presenting symptoms, life history, and present situation; a social worker, conducting an interpersonal relationships interview (14); and a psychologist administering a Rorschach and TAT. The raters were asked to check each item which represented an area of discomfort for a given patient, and then to rate the degree of disturbance in this particular area on a three-point scale. The instructions to the raters spelled out a specific procedure designed to insure that each rating would be based on a

[2] The writers are greatly indebted to David Rosenthal for his extensive help in the construction of the Q sorts.

[3] The test-retest reliability within a month for these rank orderings was computed for three randomly selected patients. The rho correlations for the Perceived Self Q sort were .93, .84, and .94 respectively; for the Ideal Self Q sort, .84, .84, and 1.00 respectively.

consideration and subjective weighting of all of the patient's significant interpersonal relationships on which information was available. After each staff member had completed his independent ratings, the three met together to discuss their ratings and to check interpretations derived from their separate methods against new evidence. This meeting yielded a common staff rating for each item, which was not a simple average of the three independent ratings, but often a new rating based on the pooled information from all three sources.[4] The common staff ratings on all items were summed, thus yielding a total score on discomfort. Separate scores were obtained on the basis of pre-therapy tests and interviews and again on the basis of post-therapy tests and interviews. The differences between the "before" and "after" scores constitute the measure of change in degree of discomfort.[5]

Measures of Effectiveness

Sociometric ratings by group. It is commonly assumed that the therapy group gives the patient the opportunity to try out new responses and, in an environment relatively free of real life threats, to develop and test new social skills which he can transfer to his everyday relationships (3, 5). To the extent to which this is true, one would expect that the patient's increased effectiveness in the group would eventually be reflected in his interpersonal relationships outside the group. Accordingly, one of our measures of effectiveness was the average evaluation of each patient by his fellow group members on a modified sociometric questionnaire. Each patient was asked to rate all other patients in his therapy group on three dimensions: the extent to which he respected them, regarded them as leaders, and desired to be friends with them. The ratings were made on a four-point scale, and were assigned scores from zero to three. To arrive at a measure of effectiveness for a given patient, the average rating he received on each dimension was computed and the three averages were then summed. The difference between the score thus obtained after five weeks of therapy and the score obtained after 20 weeks constitutes the measure of change in effectiveness. It must be stressed that this measure is based on changes in effectiveness *in the therapy group* rather than in the community. Increased effectiveness in the group is not necessarily accompanied by corresponding changes in daily life. Conversely, for some patients, behavior in the community may become more effective and yet behavior in the group may remain unchanged or even deteriorate. To obtain more direct measures of effectiveness in the community would, of course, require different sources of judgment and possibly different instruments (10).

Ineffective behavior evaluation scale. This scale consisted of sixteen items, describing kinds of difficulties a patient might experience in achieving goals that are important to him and in carrying out the social roles which he is expected to assume, e.g., difficulties due to submissiveness, indecisiveness, inability to control feelings, inadequate use of capacities, etc. The three staff members who filled out the discomfort evaluation scale also completed this form, checking each item which represented an area of difficulty for a given patient and then rating on a three-point scale the degree to which this difficulty is characteristic of the patient. The procedures for making the ratings, for arriving at common staff ratings, and for developing a total score, were the same as those used for the discomfort evaluation scale.

Measures of Self-awareness

Self-awareness Q sort. Self-awareness was defined as the extent to which a patient sees himself as others see him and was measured by the degree of congruence between the patient's perceived self and his behavior as described by trained observers. To determine a patient's degree of self-awareness, the Q sort items described earlier were used. The Perceived Self Q sort was the same as the one used for measuring self-satisfaction. The Observed Behavior Q sort was obtained by asking observers to sort the 60 items for each patient in such a way as to describe his behavior in the group.[6] The observers were two clinical psychologists and one social worker, each of whom was assigned to one of the three groups. Each had considerable training in group observation. Moreover, they were completely outside the therapy group and played no role in the therapeutic process. For these reasons, it was assumed that their perceptions of the patients' behavior in the group were relatively "objective." For each sorting, the six drives were rank ordered and the rho correlations between the Perceived Self and the Observed Behavior Q sorts were computed. The difference between the rho obtained after four weeks of therapy and the rho obtained after 20 weeks constitutes the measure of change in degree of self-awareness. A serious limitation on this measure is the fact that the observers had seen the patients outside the therapy group in order either to test or interview them; also, they observed the groups throughout

[4] Since it was suspected that the psychiatrist might be exercising a greater influence on the final staff rating than the psychologist or social worker, a comparison was made between the final staff ratings and each of the three independent ratings. This analysis showed that there was no overall tendency to adopt the ratings of any one of the three. It can be concluded, therefore, that the staff ratings do in fact represent a joint product.

[5] It was not possible for the staff members to fill out the pre-therapy evaluation scale and to hold the joint staff sessions until after the post-therapy tests and interviews had been administered. Although the ratings were based on the records of the *pre*-therapy tests and interviews, it is quite likely that they were affected by the post-therapy tests and interviews. The before- and after-scores derived from the evaluation scale are, therefore, not completely independent. A cautious interpretation of the change scores would be that they are essentially *ratings of change* rather than measures of change based on the patients' status before and after therapy. The same limitation holds true for the ineffectiveness evaluation scale described here.

[6] The test-retest reliability was obtained by asking each observer to repeat the sorting on a randomly selected patient of his group. In each case, the rho correlation was 1.00.

TABLE 1
COMPARISON BETWEEN MEAN BEFORE-SCORES AND MEAN AFTER-SCORES ON DIFFERENT MEASURES OF IMPROVEMENT (N = 15)

Measure	Lowest Score Possible*	Highest Score Possible*	Mean Before-score	Mean After-score	p-Value for Difference**
Measures of comfort					
(1) Self-satisfaction Q sort	−1.00	1.00	.44	.41	—
(2) Symptom-disability checklist	123	0	44.13	36.33	$<.01$
(3) Discomfort evaluation scale	30	0	22.14	17.20	$<.01$
Measures of effectiveness					
(1) Sociometric ratings by group	0	9	5.47	5.21	—
(2) Ineffective behavior evaluation scale	48	0	16.27	13.60	$<.01$
Measures of self-awareness					
(1) Self-awareness Q sort	−1.00	1.00	.43	.44	—
(2) Sociometric ratings predicted for self	9	0	2.88	2.63	—

* By "lowest score possible" is meant the extreme of the possible range of scores on the unfavorable end, i.e., the end denoting "poor mental health." Movement away from the lowest score possible in the direction of the highest score possible would be considered "improvement."

** The p values were obtained by using Wilcoxon's (15) nonparametric method for determining the significance of the difference between paired replicates.

the 20 weeks, so their final sortings were inevitably based on a larger sample of behavior than the initial sortings.

Sociometric ratings predicted for self. In filling out the sociometric questionnaire described above, patients were also asked to predict the ratings they would receive on each dimension from each of their fellow group members. For each dimension, the average discrepancy between the ratings the patient expected and the ratings he actually received was computed. The three averages were summed and the score thus obtained was used as a measure of another aspect of self-awareness: the patient's awareness of the kind of impression that he makes on others. The difference between the discrepancy score obtained after five weeks of therapy and the score obtained after 20 weeks constitutes the measure of change in accuracy of prediction.[7]

The seven measures can be grouped into "indirect" and "direct" measures of improvement. Those derived from the evaluation scales are indirect in that they are based on staff ratings. Those derived from the Q sort, the sociometric, and the symptom checklist are direct in that they are based essentially on the patient's "performance" in relevant test situations. Of the direct measures, those derived from the Q sort and the sociometeric have as their referent the therapy group; they measure the patient's satisfaction with, effectiveness in, and awareness of, his group behavior. One cannot assume, of course, that behavior in the therapy group is identical with behavior in the community, especially during the early stages of therapy. Performance in the group, however, can serve as a reasonable index of changes in the patient, particularly since the focus of interest here is to study interrelations among different measures rather than to evaluate a given form of therapy.

Three additional measures, which might help in providing dynamic explanations of patient behavior, were obtained. Each of the staff independently rated each of the patients with regard to (*a*) the degree of increase or decrease of anxiety displayed by the patient in the course of therapy, (*b*) the degree to which the patient tended to accept or reject repressive solutions to his problems, and (*c*) the degree to which the patient was self-revealing and discussed personal material.

RESULTS

Table 1 summarizes the data on the amount of change that took place between the before-test and the after-test on each measure. It should be noted that these data do not constitute a proper basis for evaluating group therapy since no control group was used. They are presented here in order to provide a basis for interpreting the interrelations among changes found on the different measures.

Three of the seven measures show significant change for the group as a whole: the symptom-disability checklist, the discomfort evaluation scale, and the ineffective behavior evaluation scale. It is interesting to note that these are the three measures which are based on some kind of rating of well-being by either the patient or the staff. The lack of significant

[7] A third measure of self-awareness, a distortion evaluation scale (containing four items), was also used. The range of change scores which this measure yielded was extremely small, however; 0, 1, or 2 were the only scores represented. Since it would be almost impossible to evaluate and interpret correlations based on such a distribution, it was decided to omit this measure.

TABLE 2
INTERRELATIONS AMONG CHANGE SCORES ON SEVEN MEASURES OF IMPROVEMENT (N = 15)*

	Self-satisfaction Q Sort	Symptom-disability Checklist	Discomfort Evaluation Scale	Sociometric Ratings by Group	Ineffective Behavior Evaluation scale	Self-awareness Q Sort
Symptom-disability checklist	.14					
Discomfort evaluation scale	.16	.44				
Sociometric ratings by group	−.36	−.10	.07			
Ineffective behavior evaluation scale	.02	.36	.44	−.08		
Self-awareness Q sort	.11	.54**	.18	−.31	.44	
Sociometric ratings predicted for self	.10	.14	.03	−.29	−.30	.02

* In the computation of Pearson r's, the scores on each measure were so arranged that all would move in the same direction, i.e., the higher the score, the greater the improvement. A positive correlation would mean, therefore, that individuals who improved according to one measure also tended to improve according to the other.

** $p < .05$.

over-all changes on the other measures does not necessarily mean that no changes occurred. A review of the data shows that some patients changed in a positive direction and others in a negative direction, and that these two tendencies tended to cancel each other out. An example of this type of variability is found in an analysis of patients who showed an increase and those who showed a decrease in self-awareness. When the self-awareness Q sort data were compared with the staff ratings on changes in patient anxiety level, it was found that patients whose anxiety decreased in the course of therapy became more aware of their own behavior, whereas those whose anxiety increased became less aware.

Table 2 presents the correlations between change scores on all pairs of measures. It can be seen that of the 21 correlation coefficients listed in the table, only one reaches significance at the .05 level, the r between the self-awareness Q sort and the symptom checklist. There are six negative (though nonsignificant) correlations, five of them involving the sociometric ratings by the group.

This over-all picture indicates that, at least as far as the present measures are concerned, there is no support for the assumption that changes on different measures go together and that improvement can be considered a unitary process. Even the four measures that refer specifically to the patient's behavior in the therapy group—those based on the Q sort and the sociometric—reflect this independence among criteria despite the fact that these four measures are all based on the same limited range of group behavior which maximizes the likelihood of significant correlations. In short, these data underscore the necessity for specifying the criteria and measures used in evaluation whenever results of psychotherapy are reported.

The single significant correlation that was obtained obviously has to be viewed with caution, since it could easily be a chance phenomenon. It is tempting to speculate about its meaning, however. One hypothesis that might explain the correlation is that both the symptom checklist and the self-awareness Q sort are responsive to reduction in anxiety. To the extent to which the patient's anxiety decreases, his symptoms will tend to disappear. Also, his self-awareness will tend to increase, since there will be less of a need for a defensive self-concept. Some suggestive evidence for this hypothesis was obtained from the staff ratings of changes in patient anxiety: those patients who were rated as "decreased in anxiety" showed the greatest change both on the symptom checklist and on the self-awareness Q sort.

DISCUSSION

The findings of the present study seem to provide no evidence for the assumption that changes on different measures go together. The over-all picture presented in Table 2 makes it impossible to reject the hypothesis of "no relationship." The status of this finding, however, must remain ambiguous because of questions of reliability and of inter-

pretation which cannot be answered by the present data.

First, there is some ambiguity regarding reliability. There are at least two statistical reasons why the finding of "no relationship" might fail to show up in a replication of this study. It may be that positive though small relationships exist, but that a larger sample is necessary in order to demonstrate their significance. Furthermore, on four of the seven measures there was no significant change for the group as a whole over the period of therapy (see Table 1). While it is possible, of course, that some of the patients showed real changes even if the group as a whole did not, it is also possible that the obtained changes merely represent measurement error. If the latter is true, then one can conclude very little from a lack of relationship between measures. There is, after all, no reason to expect a correlation between two sets of random errors. It can be argued that if real changes had occurred, they might very well have been related to each other. Where there was no significant change to begin with, measurement error cannot be rejected as a possible explanation.

In view of these limitations, one can only draw the modest conclusion that the data do not confirm and hence throw doubt on the frequent assumption that improvement is a unitary process.

Secondly, even if there were no question about the reliability of the finding of "no relationship," the interpretation of this finding would still be ambiguous. An over-all lack of relationship among measures would be found if the measures were completely independent. It would also be found, however, if the measures were related to each other in a complex way. That is, if under some conditions (or for certain types of patients) two measures are positively correlated and under other conditions they are negatively correlated, then an over-all correlation of around zero is likely to result.

The second interpretation provides a model for analysis and hypothesis building which is probably more meaningful from a clinical point of view. It is hard to believe that changes in self-awareness, for example, will be completely unrelated to changes in effectiveness. On the other hand, it is quite reasonable to expect that for some patients increased self-awareness will go with increased effectiveness, while for other patients it will go with reduced effectiveness. Which interpretation is correct has to remain ambiguous for the moment. Either interpretation, however, points to the need for paying attention to the specific criteria and measures that are used in evaluation. At the very least, it seems likely that not *all* measures of improvement go together or—if the second interpretation is correct—that not all measures go together for all patients and under all conditions. Further research would have to indicate what the relevant conditions are. Such knowledge would help not only in the selection of appropriate therapeutic programs for particular kinds of patients, but also in the understanding of the therapeutic process.

The present study has suggested some hypotheses about the conditions under which positive correlations between changes on two measures would be expected, and the conditions under which negative correlations would be expected. In general, it would appear that the kinds of changes that occur in group therapy can be related to the activities and experiences of the patient in the group. The correlations between changes, therefore, can also be understood in terms of these activities and experiences.

Consider, for example, the relationship between changes in symptoms and changes in self-satisfaction. One may hypothesize that for "under-controllers," i.e., patients whose ego-control is overly flexible, who tend toward variable and impulsive behavior, and who are quite conscious of their emotional difficulties and their low self-satisfaction (2), the correlation would be positive. As these patients gain greater control over their impulses and learn to stabilize their behavior, they will become more satisfied with themselves and also lose their presenting symptoms. For the "over-controllers," on the other hand, i.e., patients whose ego-controls are rigid, who deal with their difficulties in a defensive and repressive manner, and who deny that there is a discrepancy between their perceived self and ideal self, a negative correlation may be hypothesized. These patients will tend to lose their presenting symptoms only as they learn to relax their controls and to adapt to changing conditions. To the extent to which they relax their controls, however, their defensive self-satisfaction will tend to diminish. Thus,

as the patient improves symptomatically, his self-satisfaction may go down. Suggestive evidence for this hypothesis was obtained when staff members rated each patient in terms of his tendency to be "self-revealing" (typical of the "under-controller") as against a tendency to "hold back" (typical of the "over-controller"). For the patients who were rated as self-revealing, there was a significantly higher correlation between the two measures than for the patients who held back.

Aside from personality factors, such as ego-control, the nature of the patient's group experience is likely to affect changes and hence correlations between changes. The correlation between changes in self-awareness and changes in self-satisfaction, for example, may be positive if the patient is given warmth and support in the therapy group. For such a patient "to become more objective about his own behavior" may mean to learn (perhaps for the first time in his life) that others really like him. As his self-awareness increases, therefore, he would also become more satisfied with his interpersonal behavior. On the other hand, if the patient is criticized and rejected by other members of the group, one would expect a negative correlation between the two measures. For such a patient, "to become more objective" may mean to learn how unacceptable to others he really is. Thus, as his self-awareness increases, he may become more dissatisfied with himself.

The above example points to the possible negative or "anti-therapeutic" effects that group therapy may have. These possibilities become particularly apparent when the patient's effectiveness in the group (defined by sociometric ratings in the present study) is used as a measure of improvement. A great deal depends on who the other group members are and what kinds of group standards have developed. For example, in some groups a patient who has insight into his own behavior may alienate other patients by threatening their defenses.[8] Here one would expect a negative correlation between changes in self-awareness and changes in effectiveness in the group. In those groups, however, in which insight is encouraged, a positive correlation would be anticipated.

To take another example, one might hypothesize that increases in self-satisfaction depend on the extent to which the patient engages in frank interaction with other patients and expresses his true feelings. In therapy groups in which this kind of free expression is generally accepted, it would also be expected to lead to increased effectiveness in the group. In groups in which such expression is not sanctioned, however, it may antagonize the other patients and hence lead to reduced effectiveness in the group.

Finally, the group norms may be such as to accelerate the drive towards improvement. A patient who improves may receive the support and encouragement of his fellow patients, who in their turn are encouraged by his progress. Where that is true, a positive correlation would be expected between changes in symptoms and in group effectiveness. It is also possible, however, that patients mutually reinforce their defenses and develop group norms favoring a "slow-down" in improvement.[9] The more disturbed a patient is, the higher may be his status in the group. The patient who loses his symptoms may also lose his status; he may constitute a threat to the others and therefore be rejected by them.

As can be seen from Table 2, there is a slight tendency for the change scores on sociometric effectiveness to be negatively correlated with the other measures. This would suggest that some of the antitherapeutic agents discussed may have been present in the groups studied, at least during the early phases of therapy. It is important to keep these possibilities in mind; one cannot assume that patients in a group will always be accepting and supporting of one another. Such negative possibilities exist in all forms of therapy, and it is important to identify them in order to make possible a rational choice between different forms of therapy for a given patient.

SUMMARY

The purpose of the present study was to investigate the interrelations among seven meas-

[8] This possibility is suggested in a study by Norman (8), using a non-patient population. He found that individuals who had a high degree of self-insight were less accepted than those with a medium degree of self-insight.

[9] In this respect, it is interesting to compare the therapy group with the industrial work group. Numerous studies have shown that group standards can have the effect of increasing or decreasing an individual's output.

ures of improvement in therapy, representing three separate criteria of improvement: comfort, effectiveness, and self-awareness. Comfort was measured by a self-satisfaction Q sort, a symptom-disability checklist, and a discomfort evaluation scale. Effectiveness was measured by sociometric ratings and an ineffective behavior evaluation scale. Self-awareness was measured by a Q sort and predictions of sociometric ratings. Scores on these seven measures were obtained for fifteen neurotic patients at the beginning of a course of group therapy and again after 20 weeks of therapy. The change between before- and after-scores in each case constituted the measure of improvement.

Significant change was found on three measures: the symptom checklist and the two evaluation scales. Of the 21 intercorrelations between change scores, only one was significant at the .05 level. It was concluded that, at least as far as the present measures are concerned, there is no support for the frequent assumption that changes on different measures and criteria go together and that improvement can be considered a unitary process. The need for specifying the particular criterion and measure used for evaluation of psychotherapy was stressed.

Some of the ambiguities regarding reliability and interpretation of the findings were discussed. It was suggested that some of the near-zero correlations found may be due to the fact that under some conditions (or for certain types of patients) two measures may be positively correlated, while under other conditions these same measures may be negatively correlated. Some hypotheses about the conditions that may affect the direction of correlations were offered. They included hypotheses about the role of personality factors, such as ego-control, and the role of group experiences, such as the amount of support or rejection a patient receives from others and the kind of group standards that evolve.

REFERENCES

1. BION, W. R. Experiences in groups. *Hum. Relat.*, 1948, **1**, 314–320, 487–496; 1949, **2**, 13–32, 295–304; 1950, **3**, 3–14.
2. BLOCK, J., & THOMAS, H. Is satisfaction with self a measure of adjustment? *J. abnorm. soc. Psychol.*, 1955, **51**, 254–259.
3. FRANK, J. D. Group methods in psychotherapy. *J. soc. Issues*, 1952, **8** (2), 35–44.
4. HANLON, T. E., HOFSTAETTER, P. R., & O'CONNOR, J. P. Congruence of self and ideal self in relation to personality adjustment. *J. consult. Psychol.*, 1954, **18**, 215–218.
5. KELMAN, H. C. Two phases of behavior change. *J. soc. Issues*, 1952, **8** (2), 81–88.
6. KOGAN, L. S., & HUNT, J. MCV. After comment. *Psychol. Serv. Center J.*, 1950, **2**, 132–138.
7. MILLER, J. G. Objective methods of evaluating process and outcome in psychotherapy. *Amer. J. Psychiat.*, 1951, **108**, 258–263.
8. NORMAN, R. D. The interrelationships among acceptance-rejection, self-other identity, insight into self, and realistic perception of others. *J. soc. Psychol.*, 1953, **37**, 205–235.
9. PARLOFF, M. B. Some factors affecting the quality of therapeutic relationships. *J. abnorm. soc. Psychol.*, 1956, **52**, 5–10.
10. PARLOFF, M. B., KELMAN, H. C., & FRANK, J. D. Comfort, effectiveness, and self-awareness as criteria of improvement in psychotherapy. *Amer. J. Psychiat.*, 1954, **111**, 343–351.
11. ROGERS, C. R. Studies in client-centered therapy: III. The case of Mrs. Oak—a research analysis. *Psychol. Serv. Center J.*, 1951, **3**, 47–165.
12. ROSENTHAL, D. *Perception of some personality characteristics in members of a small group.* Unpublished doctor's dissertation, Univer. of Chicago, 1952.
13. STOCK, DOROTHY. *The relation between the sociometric structure of the group and certain personality characteristics of the individual.* Unpublished doctor's dissertation, Univer. of Chicago, 1952.
14. VARON, EDITH, & MAAS, H. S. Evaluating social adjustment through a systematic study of interpersonal relationships. *J. psychiat. soc. Wk.*, 1951, **21**, 3–12.
15. WILCOXON, F. *Some rapid approximate statistical procedures.* New York: American Cyanamid Company, 1949.
16. *Cornell Medical Index Health Questionnaire.* New York: Cornell Univer. Medical College, 1949.

Received July 19, 1956.

DETERMINANTS OF CONFORMITY AND ATTITUDE CHANGE[1]

KENNETH R. HARDY
Brigham Young University

An adequate understanding of human social behavior requires the simultaneous study of the nature of social influence and the causes of individual variations in susceptibility to that influence. A notable example of the recent trend to study these concomitantly is seen in the work of Asch on independence and submission to group pressure (2, Ch. 16). The study reported here had two objectives: First, to extend the work of Asch from perceptual to attitudinal judgments, from public conformity to attitude change as well. Specifically, this objective was to compare the effects upon conformity and attitude change of unanimous group pressure as contrasted with near-unanimous group pressure. The second objective was to investigate the role of affiliative motivation in affecting susceptibility to group normative pressure.

Attitude change is here considered to include any change in the organization of "motivational, emotional, perceptual, and cognitive processes with respect to some aspect of the individual's world" (10, p. 152). Attitude change arises from changes in perceived social reality, or from other influences which are thought to reflect "objective" reality dependably, such as the opinions of experts or other prestigeful persons, the evidence of research, personal experience, and the like. The importance of "social reality" (6, 16), defined in terms of the shared consensus of (significant) people in the person's environment, is now widely acknowledged as a determinant of social attitudes.

Conformity is here defined as the public avowal of a belief or attitude at variance with one's prior position, which avowal tends to correspond to the position approved by the group in which the avowal occurs. It may occur as an accompaniment of attitude change based upon a shift in perceived social reality. It may also occur in the absence of attitude change in situations where it is believed to be instrumental to the satisfaction of such motives as sex, achievement, affiliation, power, or hunger.

The type of social situation under investigation is one in which a small group of men meet informally. They state their opinions on a given topic, expressing unity of opinion. What will a person do whose position on the issue has hitherto been at variance with that of the others? Will he change his mind about the issue, or will he, at the least, publicly express agreement with the group? What will he do if another person in the group expresses a similarly deviant position? The predictions investigated were as follows: First, there would be (*a*) greater conformity and (*b*) greater attitude change under conditions of unanimous opposition to the subject's prior position, than where the subject has a "partner" who shares his opinion. This is presumed because in the "partner" situation the pressures toward conformity are less, and the criterion of social reality loses much of its force when there is a difference of opinion expressed within the group. Secondly, there would be (*a*) greater conformity and (*b*) greater attitude change by those who have greater affiliative motivation than by those with lesser affiliative motivation. This prediction is based partly on the greater desire of the former to be accepted and well thought of, combined with the common realization of the tendency of groups to reject those deviant from group norms (17), and partly on their presumably greater intellectual dependence upon social reality.

METHOD

The fundamental plan of the research was to submit a subject to a social-influence situation wherein his own attitude would be opposed by all, or alternatively, all but one of the other persons present. The focus of study was the *S*'s public response to this situation—and his private (attitudinal) response.

[1] This report is based upon a doctoral dissertation, University of Michigan, 1954, which is on file with University Microfilms, Ann Arbor, Michigan, as publication No. 8311 (price $1.56), and which contains the raw data, instruments used, and a fuller discussion of theory, method, and results.

The author wishes to express his appreciation to Dr. Daniel Katz and the other members of his committee for their counsel and encouragement. He is also grateful to Dr. J. W. Atkinson, Dr. J. D. Birch, Joseph Veroff, and Robert Hamblin for assistance in special phases of the research.

In order to accomplish this objective, a three-stage design—of pretest, experimental manipulation, posttest—was used. Male *Ss* were recruited from classes in introductory psychology and general chemistry at the University of Michigan. They were asked to volunteer two hours of their time to a study of the relative efficacy of various methods of measuring attitudes. The attitude to be studied was that toward divorce, and considerations of control and expedience led to the elimination of all Catholic, married, and Negro volunteers.

Pretest. The pretest, administered to small groups of *Ss* at a time, consisted of four instruments (discussed below), all ostensibly representing methods of measuring attitudes. Actually, the major purpose of the pretest was to secure data on the *Ss*' affiliative motivation and attitudinal position respecting divorce so that they could be assigned to the appropriate experimental variation.

The first instrument given was the TAT-type measure of affiliative motivation (n Affil) developed by Atkinson, Heyns, Shipley, and Veroff (3, 19). The standard story blanks were distributed and standard instructions for administering a projective measure of motivation were then read (cf. 13). Four pictures were shown in the following sequence: (*a*) two men conversing; (*b*) group of seven men around a table; (*c*) older woman seated, looking away from young man by window; (*d*) man by lamppost (TAT 20).

It was intended to structure a situation that would be only mildly arousing of affiliative motivation, so that affiliation imagery would be aroused mostly by the pictures shown. The small groups tested consisted of strangers or relatively casual acquaintances meeting in a seminar room on campus. The testing was usually preceded by a minor amount of "small talk" as the men arrived in the room. This situation thus appeared to be about as affiliatively arousing as the "neutral" condition of Shipley and Veroff (19), a conjecture that is substantiated by a comparison of scores obtained in the present study with those obtained on a similar population by these investigators.

Following the completion of the stories, *Ss* were asked to write their names on the back of each sheet. Assurances were given of the confidential treatment of this and other material collected during the study. The blanks were then collected, and a questionnaire booklet containing the other three instruments was distributed.

The first questionnaire was the California F scale (1, pp. 255–257), introduced as dealing with general public opinion. This scale, and the self-ratings discussed below, were primarily used to bolster the facade that the study was investigating various methods of measuring attitudes, at the same time drawing attention away from the divorce issue. They were also used to explore possible relationships they might have with the experimental variables.[2]

The second questionnaire, which pertained to the divorce issue, was introduced as dealing with a specific attitude quite intensively rather than dealing superficially but extensively with many attitudes as the preceding questionnaire had done. The first page asked the *S* to write "your attitude and feelings toward divorce and the reasons you feel the way you do about it." He was allowed about ten minutes to do this. The objective of this section was to provide arguments that his supporter could use in the group situation. The next section consisted of three parts. The first part was a nine-point graphic scale of leniency toward divorce, on which *S* indicated his own position and those of his parents and friends. This scale was also used in the group-influence situation, where *S* had to state his position publicly. A comparison of the two positions taken by *S* provided the operational definition of conformity. The second part consisted of 12 attitude statements dealing with various aspects of divorce: the nature of the marital contract, divorce as a solution to an unhappy marriage, effect of divorce upon children and society generally, the degree of present abuse of divorce, and the obligation partners should feel to remain married. There were six "liberal" and six coordinate "conservative" statements. This set of statements, also given after the group situation, provided the operational definition of attitude change. For example, if an *S* marked "−" (mildly disagree) to a given statement initially, but afterward marked "+" (mildly agree), this would be scored as a change of two units. If he marked "++" (strongly agree) initially and "−" later, three units of change would be recorded. If the change occurred in the direction of the group influence, it was scored positively, and if in the opposite direction, scored negatively. The algebraic sum of these shifts over the twelve statements was the *S*'s attitude-change score. The third part of this questionnaire contained two questions that assessed the strength and certainty of the *S*'s attitude toward divorce.

The final instrument in the questionnaire booklet was a series of ratings for 23 personality traits. Most of the traits were adapted from Fiske (7), who in turn adapted them from a longer list used by Cattell (4). In Fiske's factor-analytic study, seven of the self-rated traits had loadings of .40 or better on a centroid factor he labeled "conformity."

Experimental manipulation. The group-influence situation came after an interval of four to eleven weeks following the pretest. In the interim, the n-Affil scores were computed and the resulting distribution divided into high and low n-Affil groups.[3] These groups were then assigned to the *partner* or *no-support* variations so as to control for initial position on the divorce issue (as indicated by the graphic scale) and time elapsing since the pretest. Each *S* met with a group of six other students, all of whom were presumably *Ss*. Actually these six were confederates who had been instructed to give standardized "arguments" using five kinds of appeals: emotional, personal contact or experience, scientific evidence, appeal to authority, and acceptability of change as a result of further thought and

[2] Analysis of results showed no relationships between the F scale and either conformity or attitude change. Negative results were also found between the "conformity" factor of the self-ratings and these dependent variables.

[3] Reliability was established by comparing the author's scoring of two samples of stories obtained by previous research with the pooled scoring of the authors of the n-Affil manual (7) on these same stories. The average rho for the two samples was .80 for total scores.

investigation. The content of each argument centered around some aspect of divorce sampled by the attitude statements in the questionnaire.

As an example, the antidivorce argument appealing to authority and dealing with the effectiveness of divorce as a solution to an unhappy marriage is as follows: "Some people say that a couple ought to get divorced if they are incompatible. A psychiatrist who is a friend of the family told me that the real reason marriages are unhappy is that the couples have not really learned how to solve their problems. The definition of a happy marriage is not one that has no differences or problems; it is one where the husband and wife know how to solve their problems as they arise. Getting divorced and later remarrying doesn't help people to learn these techniques. What they need to be happy is to learn how to get along together." The confederate was free to state the argument in his own words.

It should be noted that this technique of not only stating a judgment but also buttressing it with "reasons" represents a methodological innovation, one that represents the ordinary social-influence situation more faithfully than the perceptual situations studied by Asch (2) and Sherif (18), in which the judgment is immediately given and there is no room for discussion.

In the group situation, *S*s were seated in a semicircle (the confederates were free to take any seat) facing the *E*, who began by stating,

> During the first hour we worked with several methods of measuring attitudes. Today we are going to work with another method which I might call a "group interview." What I shall do is call on each of you in turn and ask you to tell me how you feel about a particular topic and why you feel the way you do about it. The topic is divorce. I shall write down in summary form what you say, and then attempt to restate it to see if I have it accurately.

After each interview was completed, *E* held up a large placard on which was the graphic scale used during the pretest. After the person stated his position, together with any reasons for placing himself there, *E* called on the next individual in his list of names. This procedure was repeated until all the confederates, and finally the *S*, had spoken.

In the no-support variation, all of the confederates opposed the *S*'s previously indicated position. Since two sets of arguments had been constructed, one favorable, the other unfavorable to divorce, the set used was the one opposed to *S*'s position as indicated on the pretest. A position on the graphic scale paralleled each argument. The arguments were always given in the same sequence. In each case the fifth man "agreed" with the others but added no arguments to those already given. In the partner variation, the fifth man opposed the others, giving arguments the *S* had used during the pretest; he placed himself at the *S*'s initial position on the graphic scale. By this procedure the amount of new ideas presented to the *S* was kept constant across the variations. The fifth man was used so that the group norm would be clearly established by the first four men and maintained by the sixth person.

Upon completion of the "group interview," *E* said, "Now there is one other thing I'd like you men to do before we are through and that is to fill out a short questionnaire for me individually. In order to do this, I will assign you to separate rooms." Each man was handed a (posttest) questionnaire as he went to his room, but the *S*, of course, was the only one who filled it out.

Posttest. This questionnaire, which was filled out privately, contained measures of liking of the group, perception of the group situation, and repeat measures of the attitude toward divorce. Upon completion of the questionnaire, *E* interviewed the *S* and explained to him the real nature of the study. At this time *S* was given the opportunity to serve as a confederate (which most *S*s were eager to do, especially those most affected by the group situation). The major source of confederates, however, was volunteers from classes in social psychology and human relations.

Although a total of 45 "group interviews" were conducted, in 13 cases the experimental conditions were not met by the failure of one or two confederates to appear, and in one case there were two supporters due to a mix-up in instructions. The results are based upon the 31 remaining cases.

RESULTS

The first question to be answered is, "Did the situation really provoke conformity and changes in attitude?" With regard to conformity, 14 men (45%) shifted two or more scale points on the graphic scale, in the direction of the group influence. (The average distance from initial position to the group norm was 4.36 scale points.) In contrast, only three men shifted in the opposite direction, and none of them changed more than one point. Similarly, 14 men had a positive attitude-change score of six or more points; that is, they changed their position on at least half of the attitude statements (or the equivalent). Only two men had a negative score this high. It should be noted here that the correlation between conformity and attitude change was .51. Observations of these men confirm those of Asch in detecting more or less severe signs of emotional activation in the group situation. One *S*, for example, spoke very rapidly (atypical of him) in pronouncing his agreement with the others, then blurted, "Boy, it's hot in here; can't we open a window?" He was the only one to notice the "heat"!

Effect of social support. Results of *t* tests indicated that the absence of social support was significantly related to attitude change, but the relationship to conformity fell short of significance, as shown in Table 1. The hypothesis regarding the effect of social support upon attitude change thus received confirma-

TABLE 1

EFFECT OF SOCIAL SUPPORT ON CONFORMITY AND ATTITUDE CHANGE

Dependent Variable	Condition			
	No support ($N = 16$)	Partner ($N = 15$)	t	p
Conformity				
Mean	1.81	1.16	1.30	.11
SD	1.55	1.12		
Attitude change				
Mean	6.97	2.00	1.79*	$< .05$
SD	9.44	5.05		

* A formula recommended by Dixon and Massey (5, pp. 104–105) was used here because of heterogeneity of variance between the two groups.

TABLE 2

EFFECT OF n AFFILIATION AND SUPPORT ON CONFORMITY

a. Conformity in Variations of n Affiliation and Support

Degree of support	n Affiliation		
	High	Medium	Low
No support			
Mean	1.97	2.97	−.175
SD	1.41	.58	.43
N	6	6	4
Partner			
Mean	.06	1.83	1.58
SD	.76	1.15	.59
N	5	5	5

b. Results of Variance Analysis

Source	*df*	Mean Square	*F*
Support	1	1.51	1.49
n Affiliation	2	9.88	9.77
Interaction	2	20.85	20.62
Within cells	25	1.01	

$F_{.01} = 5.57$ for 2 and 25 *df*

tion, but the evidence is ambiguous respecting the effect of support upon conformity.

Effect of affiliative motivation. A comparison of high and low n-Affil groups showed no differences between these groups on either conformity or attitude change. However, it seemed possible that social support might be interacting with n Affil in such a way as to cancel out differences between the affiliation groups, or that an alternative division of the n-Affil distribution might show positive relationships.

This second possibility recommended itself from some research done on n Achievement

TABLE 3

EFFECT OF n AFFILIATION AND SUPPORT ON ATTITUDE CHANGE

a. Attitude Change in Variations of n Affiliation and Support

Degree of support	n Affiliation		
	High	Medium	Low
No support			
Mean	6.75	11.00	1.25
SD	7.37	11.37	3.36
N	6	6	4
Partner			
Mean	.40	1.20	4.40
SD	4.84	4.26	5.08
N	5	5	5

b. Results of Variance Analysis

Source	*df*	Mean Square	*F*
Support	1	153.78	2.51
n Affiliation	2	43.44	
Interaction	2	242.85	3.96
Within cells	25	61.35	

$F_{.05} = 3.38$ for 2 and 25 *df*

by Lowell (12), McClelland and Liberman (14), and Moulton (15), who found more clear-cut relationships with certain perceptual and behavioral phenomena on the basis of a high-medium-low division than with a dichotomization of the n-Ach distribution. They found the high group, which is conceived to be positively motivated toward achievement, to suffer no decrement in performance nor perceptual defense under achievement-arousing conditions. The middle n-Ach group, which is conceived to be ambivalently motivated toward achievement, *does* show defensiveness to perception of failure words, and a performance decrement under arousal conditions.

Interaction of social support and n Affil. To examine these two possibilities, analyses of variance were made, using the two categories of social support, and three categories of n Affil based upon a division of that distribution into three groups as nearly equal in size as possible. Table 2a reports the *N*s, means, and *SD*s of the six resulting cells for conformity; Table 3a the corresponding data for attitude change. Before analyses of variance were made, the cell frequency (with the accompanying sums of squares) of the "Low-n-Affil, No-support" cell was adjusted to make possible an analysis procedure based

upon proportional cell frequencies (cf. 11, pp. 151–155). Also, before each analysis, the assumption of homogeneity of variance within cells was checked by using Bartlett's test, and was verified in each case. Table 2b gives the results of the variance analysis for conformity, Table 3b for attitude change.

The results for conformity indicate that both n Affil and the interaction component are highly significant factors, but support is not. Within the table the following results were obtained:

1. The effect of n Affil is significant within each support variation. In the no-support condition, the high and medium groups are significantly more conforming than the low group. In the partner variation, however, the medium and low groups are significantly more conforming than the high group.

2. If the support variations are combined, then the medium n-Affil group is more conforming than either the high or low groups.

3. The no-support variation is more conformity-producing for the high-need group and less conformity-producing for the low-need group than the partner variation. There is no difference between the support variations for the middle-need group.

From these results it would appear that conformity is a joint function of the social conditions and the affiliative motivation of the individual. For those high in n Affil the presence of a partner results in the disappearance of conformity that is definitely present under conditions of no support. Those in the middle range of the n-Affil distribution do not drop significantly in conformity with the presence of a partner. Those in the low n-Affil group, who do not conform at all under conditions of no support, show a significant *increase* in conformity when a partner is present.

In regard to the other dependent variable, attitude change, somewhat similar fluctuations in mean values occur, but none of the factors is statistically significant. (If the between-groups variance is tested before the interaction component is accepted as significant (cf. 11, p. 166), the resulting F ratio is 2.37, which leaves the interaction of doubtful significance.)

Though there are somewhat similar patterns for conformity and attitude change, there is one important difference. The medium-need group conformed whether or not a partner was present; but they tended to change less in attitude in the partner situation than in the no-support situation (p about .05).

DISCUSSION

These results can reasonably be interpreted in a way comparable to that previously employed for achievement motivation. The person scoring high in n affiliation is conceived to be positively motivated toward affiliation. When he is unanimously opposed in attitude, he changes his attitude to join his potential friends. When this group of people is divided in attitude, the social reality of the opposing position is much less compelling, as also the perceived necessity to conform to gain social acceptance. Under these conditions, his thinking and outward actions change but little.

The person obtaining a moderate score on n Affil is conceived to be ambivalently and anxiously motivated toward affiliation. His fear of rejection leads to a more pronounced conformity. His thinking, as in the case of the positively motivated person, leans heavily on social reality, so that he shifts attitudinally in the nonsupport situation, but much less so under conditions of support. In contrast to the positively motivated person, however, his anxiety for approval tends to maintain his conformity even in the support situation, since he derives less assurance from the presence of a lone partner.

If the low-scoring person experiences a negligible arousal of affiliative motivation in the social-influence situation, then his behavior should be a function of other factors present, in particular the content of the influential communications. To the extent that he is less dependent upon social reality, he should be more free to develop beliefs based on other considerations. Hovland, Lumsdaine, and Sheffield (9) report that for those initially opposed to a point of view being advocated, greater opinion change resulted when both sides of the issue were presented (with the favored view given preponderant consideration) than when only the favored position was considered. This difference was especially pronounced for those of high school education or better. It is apparent that their situation and results are similar to those of the low-affiliation group in the present study, where greater attitude change occurred upon presentation of both sides of the issue, i.e., under conditions of support. The conform-

ity of this group is interpreted as a by-product of attitude change.

Analyses of individual cases, the posttest questionnaire and interview, and the pretest self-ratings give equivocal support to these interpretations, which of course remain highly tentative. Atkinson, in a personal communication, reports data indicating a curvilinear relationship between n Ach and n Affil, with the high n-Affil group also high on n Ach, but the medium n-Affil group low on n Ach. Since McClelland (13) found a strong negative relationship between n Ach and independence from group pressures (with subjects from Asch's studies), Atkinson believes the greater conformity of the medium n-Affil group in the present study to be due to their lower n Ach. It is probable that acceptance of social influence is mediated sometimes by the desire to be "right" or accurate (which is linked to n Ach), sometimes by the desire for social acceptance (n Affil), and sometimes by other factors. Which factors are involved will depend on the type of issue and the nature of the situation, as well as on the motivational structure of the person. The task set for Asch's subjects, for example, stressed accuracy (cf. 2, p. 452), while the present task did not.

Further research should systematically vary the nature of the task situation and the importance of the issue under study, among other variables, to clarify the relationships involved.

SUMMARY

This investigation studied the effect of affiliative motivation and social support upon conformity and attitude change. College males varying in affiliative motivation were subjected to a group-influence situation where their previously measured attitude toward divorce was either unanimously opposed, or opposed with the exception of one supporter. The subject's public reaction (conformity) was measured, and also his private response (attitude change). The results indicated conformity to be a joint function of affiliative motivation and conditions of social support, with roughly similar but less significant results for attitude change. The group with medium-need scores, interpreted to be ambivalently motivated toward social acceptance, conformed under both conditions, but changed in attitude only in the no-support situation. The high-need group, conceived to be positively motivated toward affiliation, conformed and changed in attitude under conditions of no support, but not under support conditions. The low-need group, which was least affected, changed more under conditions of support than no support, but were considered as responding more to the content of the influence situation than to its social structure.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWICK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. ASCH, S. E. *Social psychology*. New York: Prentice-Hall, 1952.
3. ATKINSON, J. W., HEYNS, R. W., & VEROFF, J. The effect of experimental arousal of the affiliation motive on thematic apperception. *J. abnorm. soc. Psychol.*, 1954, **49**, 405–410.
4. CATTELL, R. B. Confirmation and clarification of primary personality factors. *Psychometrika*, 1947, **12**, 197–220.
5. DIXON, W. J., & MASSEY, F. J., JR. *Introduction to statistical analysis*. New York: McGraw-Hill, 1951.
6. FESTINGER, L. Informal social communication. *Psychol. Rev.*, 1950, **57**, 271–282.
7. FISKE, D. W. Consistency of the factorial structure of personality ratings from different sources. *J. abnorm. soc. Psychol.*, 1949, **44**, 329–344.
8. HEYNS, R. W., VEROFF, J., & ATKINSON, J. W. A manual for scoring thematic apperception stories for the affiliation motive. Unpublished manual, Univer. of Mich., June 1953.
9. HOVLAND, C. I., LUMSDAINE, A. A., & SHEFFIELD, F. J. *Experiments on mass communication*. Princeton, N. J.: Princeton Univer. Press, 1949.
10. KRECH, D., & CRUTCHFIELD, R. *Theory and problems of social psychology*. New York: McGraw-Hill, 1948.
11. LINDQUIST, E. F. *Statistical analysis in educational research*. New York: Houghton Mifflin, 1940.
12. LOWELL, E. L. The effect of need for achievement on learning and performance. *J. Psychol.*, 1952, **33**, 31–40.
13. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
14. MCCLELLAND, D. C., & LIBERMAN, A. M. The effect of need for achievement on recognition of need-related words. *J. Pers.*, 1949, **18**, 236–251.
15. MOULTON, R. Relationship of need for achievement to perceptual sensitivity under two degrees of motive arousal. Unpublished honors thesis, Univer. of Mich., 1952.
16. NEWCOMB, T. M. Autistic hostility and social reality. *Hum. Relat.*, 1947, **1**, 69–86.
17. SCHACHTER, S. Deviation, rejection and communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 190–207.
18. SHERIF, M. A study of some social factors in perception. *Arch. Psychol.*, 1935, No. 187.
19. SHIPLEY, T. E., JR., & VEROFF, J. A projective measure of need for affiliation. *J. exp. Psychol.*, 1952, **43**, 349–356.

Received May 10, 1956.

TEST-RETEST STABILITY OF INTERACTION PATTERNS DURING INTERVIEWS CONDUCTED ONE WEEK APART[1]

GEORGE SASLOW, JOSEPH D. MATARAZZO, JEANNE S. PHILLIPS, AND RUTH G. MATARAZZO

Massachusetts General Hospital and Harvard Medical School[2]

Two previous studies (5, 6) were devoted to an investigation of the stability of interaction patterns during interviews conducted on the same day by two different interviewers. In the first study (6) of 20 subjects, we reported that interaction patterns for any given individual during a partially standardized psychiatric interview had markedly high stability coefficients across the two interviewers. In addition to demonstrating, for the interview as a whole, the stability of patient interaction patterns (effects) when the stimulus (interviewer's behavior) was controlled along certain dimensions, we observed that the highly stable interviewee patterns could be reliably modified by planned changes in the interviewer's behavior from one subperiod of the interview to another. The second study (5) was an exact replication of the first, and all the major findings were cross validated. Having thus studied stability and predictable modifiability of individual interaction patterns during two interviews conducted on the same day, we decided to examine also the stability of the interaction patterns when the test-retest interval was extended to one week. In order to control for possible interviewer interaction effects (interaction in the statistical sense) the same interviewer was used for both interviews of the present study.

The Interaction Chronograph, devised by Chapple (1, 2) for recording certain temporal aspects of verbal and gestural interaction, was the instrument of measurement used in this and the two previous studies. A more complete description of this method of observation and the variables studied will be found in an earlier report (6). A detailed review of the pertinent literature, and a history of the theory and development of the interaction chronograph, will be found in a recent article (4).

[1] This investigation was supported by a research grant (M-1107) from the National Institute of Mental Health, of the National Institutes of Health, U. S. Public Health Service.

[2] The data of this study were collected while the authors were at the Washington University School of Medicine.

Briefly, the "rules" of the standardized interview used in this and the two previous investigations require the interviewer to follow certain behavioral patterns during each of five prescribed subperiods of the interview. Periods 1, 3, and 5 consist of free give-and-take interviewing, while Periods 2 (silence) and 4 (interruption) are stress phases of the interview. The interviewer's varied behavior during these five periods constitutes a sequence of different behavioral situations, or what can be thought of as varying miniature interpersonal situations, designed to elicit the interviewee's characteristic interaction patterns. It should be noted that the stress of Periods 2 and 4 is provided by the overt behavior ("not responding" and "interrupting," respectively) of the interviewer, and not by "content" of what he says to the interviewee. The durations of Periods 1, 3, and 5 are fixed (10, 5, and 5 minutes, respectively), whereas the durations of Periods 2 and 4 are variable. The latter periods consist, respectively, of 12 failures to respond or 15 minutes, whichever is shorter; and 12 interruptions or 15 minutes, whichever is shorter.

More complete definitions of the interaction variables can be found in our earlier reports (4, 6). Briefly, they are: (*a*) *A's Units*, the frequency of the patient's actions; (*b*) *Action*, the average duration of A's actions; (*c*) *Silence*, the average duration of A's silences; (*d*) *Tempo*, the average duration of each action plus its following inaction as a single measure; (*e*) *Activity*, the average duration of each action minus its following inaction expressed as a single measure; (*f*) *A's Adjustment*, the average duration of A's "interruptions" of B minus the average duration of A's "failures to respond to" B; (*g*) *B's Adjustment*, the average duration of the other (usually the interviewer) person's adjustment; (*h*) *Initiative*, the per cent frequency with which A initiates to B following a double silence, as in Period 2; (*i*) *Dominance*, the relative frequency with which A dominates B in a double action (interruption), as in Period 4; (*j*) *A's Synchronization*, the frequency with which A fails to synchronize with B either by failing to

respond to B or by interrupting B; (*k*) *B's Synchronization*, the frequency of the other person's (usually the interviewer's) failure to synchronize with A; and (*l*) *B's Units*, the frequency of the other person's (usually the interviewer's) actions.

Method

In all aspects but two (the use of only one interviewer, and a seven-day test-retest interval) the present study is an exact replication of the two earlier experiments. Twenty additional outpatient *S*s were interviewed twice, seven days apart, by the older of the two interviewers who participated in the previous studies. The rooms and the observer were the same; the *S*s were all white; they ranged in age from 16 to 61 (median age of 34.5); and they had presenting complaints typical of the two earlier series of twenty patients each. In the first and second studies there were 11 men and 9 women, and 8 men and 12 women, respectively. The present sex distribution was 9 men and 11 women.

The present series was begun at the same point in time as the first study (6) and was completed at approximately the same point in time as the second study (5). Since the present interviewer was one of the two participants in the earlier studies, it is assumed that the amount of previous experience of the interviewer, himself, with the standardized interview is not a relevant variable in the present study.

Results

As indicated above, twenty minutes of the standardized interview are fixed and, depending upon the subject's behavior in Periods 2 and 4, up to thirty more minutes, totaling fifty in all, are possible to complete the standardized interview. Table 1 shows the mean length of the 20 pairs of interviews for both initial test and retest.

The results indicate that *both* on initial test and retest one week later the interview lasted on the average approximately 32 minutes with the third sample of subjects; a finding similar to the first and second studies. The range, from 25.3 to 41.6 minutes, is almost identical (24.5 to 41.1) with that of our second study (5). Also, as in both our earlier studies, the mean lengths of Period 2 (7.40 and 7.72 minutes) and Period 4 (2.37 and 2.34 minutes) which are free to vary were almost identical on both occasions. The test-retest value of rho (.564) for length of Period 2, as in the two earlier studies, is significant at the .01 level; whereas, once more, the extremely limited range in Period 4 makes the obtained value of rho (.180) of little practical use.

TABLE 1

Analysis of Mean Length of Total Interview in Minutes for Group of 20 Patients Interviewed Twice by the Same Interviewer

	Mean	Range
Length of all 40 interviews	31.94	25.3 to 41.6
Length of all first interviews (N = 20)	31.50	25.3 to 41.6
Length of all second interviews (N = 20)	32.40	27.1 to 40.6

Table 2 presents, for the 20 new patients, the means, standard deviations, and ranges (for the total interview) of nine of the interaction chronograph variables.

Mean values for the first and second (repeat) interview are given, with rho and Pearson coefficients of correlation (stability coefficients) shown in the second from the last and last columns, respectively. In these last two columns, rhos and Pearson coefficients of correlation (r) are reproduced from the two earlier studies for comparison. For any given variable the three values in these columns, from top to bottom, represent: the values for the present study, those for the replication study (5), and the values for the original study (6), in that order.

Whether rho or (r) is used, the values in Table 2 show for the present study that, for the interview as a whole, there is again a fairly striking stability in interaction variables for the 20 pairs of interviews separated in time by a seven-day interval. For rho, eight of the nine variables are significant at the .01 level while the remaining one reaches the .05 level. Using r, six coefficients are at the .01 level and three at the .05 level. The ranges for rho and r for these nine variables are .528* to .776**, and .464* to .890**, respectively.

When these are compared to the range of rho for the original (.668** to .874**), and replication studies (.717** to .941**), and the range of r for the original (.726** to .956**), and replication studies (.741** to .926**), it is seen that, despite their statistical significance, the obtained stability coefficients of the present study are slightly lower than those obtained in the two earlier studies with a test-retest interval of a few minutes and the use of two different interviewers.

The variables which seem to have relatively lower stability over a one-week interval are *Dr.'s Adjustment* and *Pt.'s Synchronization*,

TABLE 2

SUMMARY OF DATA FOR MAJOR I-C VARIABLES ACROSS TOTAL INTERVIEW WITH ONE DOCTOR ON TWO OCCASIONS

Variable	One Week Apart		rho[1]	r[2]
	First Interview	Second Interview		
1. Mean Pt.'s Units	72.30	78.30	.765**	.890**
SD	15.75	19.54	(.917**)	(.926**)
Range	43 to 118	48 to 131	(.807**)	(.747**)
2. Mean Pt.'s Action[3]	45.70	37.75	.688**	.464*
SD	30.58	16.68	(.914**)	(.786**)
Range	19 to 143	11 to 80	(.757**)	(.956**)
3. Mean Pt.'s Silence	9.15	9.50	.642**	.648**
SD	1.88	2.16	(.828**)	(.744**)
Range	6 to 14	7 to 14	(.731**)	(.759**)
4. Mean Pt.'s Tempo	46.00	44.40	.735**	.764**
SD	14.19	14.59	(.941**)	(.905**)
Range	27 to 96	22 to 84	(.805**)	(.908**)
5. Mean Pt.'s Activity	28.00	25.75	.738**	.768**
SD	14.19	13.60	(.910**)	(.891**)
Range	8 to 76	3 to 62	(.874**)	(.930**)
6. Mean Pt.'s Adjustment	−.89	−1.15	.776**	.757**
SD	.77	1.39	(.821**)	(.853**)
Range	+.58 to −2.50	+.89 to −5.24	(.716**)	(.802**)
7. Mean Dr.'s Adjustment	−2.09	−2.17	.558**	.554*
SD	.77	.87	(.780**)	(.749**)
Range	−.98 to −3.34	−.95 to −3.69	(.668**)	(.737**)
8. Mean Pt.'s Synchron.	.86	.88	.544**	.526*
SD	.05	.07	(.717**)	(.741**)
Range	.76 to .93	.73 to 1.01	(.691**)	(.726**)
9. Mean Dr.'s Units	59.65	66.25	.528*	.823**
SD	13.88	18.49	(.880**)	(.909**)
Range	37 to 96	37 to 118	(.782**)	(.772**)

[1,2] The three values shown in the Pearson *r* (and rho) columns for *each* variable are, from top to bottom, the *r* for the present study, the *r* for replication study, and the *r* for the original study.

[3] Values for variables 2 through 7 are given in hundredths of a minute. To convert to seconds, multiply the given value by 0.6.

* Significant at the .05 level of probability.

** Significant at the .01 level of probability.

both dropping from stability values of approximately .740** in each of the two earlier studies to approximately .540* in the present study. On the other hand, the most stable variable was again found to be *Pt.'s Units*, with a rho of .765** and an *r* of .890**.

As can be seen from the means, standard deviations, and ranges in Table 2, all the interaction variables, with the possible exception of *Dr.'s Adjustment* and *Pt.'s Synchronization*, show an absolute as well as relative stability across the two interviews for the interview as a whole. Thus, for example, the ranges for *Pt.'s Units* show that during the same standardized interview one *S* interacted 43 times during the first interview and 48 times during the second interview one week later, whereas a second *S* interacted 118 times during the first interview and 131 times during the second. Similar comparisons can be made for each of the other variables shown in Table 2, although it is apparent from the magnitude of the correlations that, for some variables, unlike the case just described for *Pt.'s Units*, the extremes of the range need not necessarily represent the same individual in every case.

In all, the data in Table 2 would thus appear to cross validate our earlier finding that, using partially standardized interview conditions, and taking the total interview as the sample, patients show similar interaction patterns from one interview to another. This finding appears to hold whether the test-retest interviews occur on the same day (as in the earlier studies) or one week apart (as in the present study). The similarity on retest is marked in all three studies, but is somewhat greater when the interviews occur only a few minutes apart, even if they are conducted independently by two different interviewers.

Table 3 presents a further cross validation of the earlier finding that these interaction patterns are invariant only as long as stimulus (interview) conditions remain the same across the two interviews for the total interview. Whether on test or retest, as the interviewer

TABLE 3

Means and Standard Deviations for Major I-C Variables for Individual Periods of Two Interviews With One Doctor on Two Occasions One Week Apart

Period		Pt.'s Units		Pt.'s Action		Pt.'s Silence		Pt.'s Tempo		Pt.'s Activity	
		First	Second	First	Second	First	Second	First	Second	First	Second
I	Mean	20.25	24.05	72.80	50.70	8.30	8.25	81.10	59.00	64.70	42.55
	SD	7.28	10.14	110.80	40.85	1.58	2.05	110.30	41.63	117.09	42.52
II	Mean	13.10	13.10	45.55	43.30	14.75	15.00	60.50	58.65	30.95	28.35
	SD	1.76	.44	37.88	25.43	5.88	5.62	40.11	26.46	36.38	25.54
III	Mean	11.00	11.95	53.10	48.85	8.85	8.90	62.20	58.40	44.20	39.90
	SD	4.72	5.22	53.74	28.56	2.71	2.78	52.81	28.84	54.77	28.11
IV	Mean	15.90	15.05	9.35	8.60	5.40	6.40	15.05	15.70	3.95	2.35
	SD	4.54	3.79	4.02	2.38	1.74	3.23	4.13	3.42	4.38	4.98
V	Mean	11.95	14.15	48.05	37.50	8.45	8.60	56.50	46.55	40.00	28.90
	SD	4.27	5.52	22.94	18.03	1.78	2.18	23.66	17.74	22.21	18.72
Total	Mean	72.30	78.30	45.70	37.75	9.15	9.50	46.00	44.40	28.00	25.75
	SD	15.75	19.54	30.58	16.68	1.88	2.16	14.19	14.59	14.19	13.60

Period		Pt.'s Adjustment		Dr.'s Adjustment		P.t's Synchronization		Dr.'s Units	
		First	Second	First	Second	First	Second	First	Second
I	Mean	−.91	−1.05	−.63	−.45	.98	.98	18.70	21.75
	SD	1.31	1.33	.75	.21	.06	.05	6.88	9.38
II	Mean	−.97	−.83	−10.87	−11.70	.29	.36	3.75	4.20
	SD	1.05	.79	3.86	3.88	.16	.21	2.16	1.99
III	Mean	−1.38	−1.53	−.67	−.70	1.02	1.01	10.60	11.35
	SD	2.03	1.85	.48	.51	.05	.03	4.60	4.77
IV	Mean	−.67	−1.04	1.22	1.39	.96	1.00	15.10	15.40
	SD	.83	2.48	.90	1.42	.08	.08	3.38	3.06
V	Mean	−.52	−1.30	−.47	−.58	1.00	.97	11.55	13.55
	SD	1.15	1.80	.22	.33	.02	.06	3.99	5.64
Total	Mean	−.89	−1.15	−2.09	−2.17	.86	.88	59.65	66.25
	SD	.77	1.39	.77	.87	.05	.07	13.88	18.49

varied his own behavior within the interview—from one subperiod to another—there is seen in Table 3 a corresponding change in the means obtained by patients with him for any one variable.

Thus, for example, during the first interviews the average interviewee values for silence in Periods 1 to 5, respectively, vary from 8.30 up to 14.75 and back down to 8.85 and down still further to 5.40 and finally up again to 8.45 hundredths of a minute as the interviewer's own silence behavior varies (see "rules" in earlier reports) from responding quickly (within one second of the patient's last utterance) in Period 1 to not responding for 15 seconds in Period 2, to responding within one second again in Period 3, to interrupting the patient in Period 4, and to responding within one second again in Period 5. Also shown in Table 3 is the almost identical pattern for the second, or retest, interview. The value for F across these 5 correlated silence means (3, pp. 284–296) was 26.06 for the first interview and 26.30 for the second interview a week later. Both of these values are significant at the .001 level of confidence.

Further examination of the data in Table 3 reveals that, as in the two earlier studies, *Pt.'s Action* is also a function of what the interviewer is doing during a particular subperiod of the interview. On the first occasion, for example, when the interviewer was conducting a free give-and-take interview in Periods 1, 3, and 5, the patients' average duration of utterance, *Pt.'s Action*, was 72.80, 53.10, and 48.05, whereas while he was interrupting them in Period 4, the patients' average action fell dramatically to a value of 9.35 hundredths of a minute. Interviewer's lack of action in Period 2 (i.e., silence stress) likewise was associated with a slight drop in patient action to an average value of 45.55. A similar pattern across the five subperiods is evident for patients' action during the repeat, or sec-

ond, interview. The values of F are 10.79 (.001 level) for the second interview and 3.22 (just short of the .01 level) for the first interview.

It should be pointed out that these subperiod means for action and silence, as well as the other variables shown in Table 3, are in reality averages of averages. That is, Chapple obtains an individual's pattern for each subperiod by dividing the obtained value for any given variable by the number of patient units in that period. For example, a patient whose units for the 5 subperiods were 20, 12, 10, 12, and 10, and whose corresponding actions were 800, 288, 400, 144, and 400, would be scored as having average subperiod actions of 40, 24, 40, 12, and 40. For any given individual, these last figures, being averages, are thus relatively independent of both the number of patient units during the 5 different subperiods, and the correlated fact that the 5 periods vary in length (as described earlier). As averages they are also relatively independent of individual differences in interaction rates (patient units) among patients and thus make possible comparisons between individuals. The means shown in Table 3 are therefore averages of these individual averages across the group of twenty patients.[3]

It is of interest to examine the records of the 20 patients individually to see whether or not the statistically significant group trends shown in Table 3 for *Action* and *Silence* are also true for individual cases. Analysis of the individual records indicates that, independent of each patient's own baseline, all 20 showed a drop in *Action* from Period 3 to Period 4 with a subsequent increase in Period 5. The pattern held individually for all 20 patients in both the first and second interviews. This finding cross validates the earlier finding of a similarly unanimous pattern (5).

As in the previous study, there is a less striking finding of increased *Silence* in Period 2 relative to Periods 1 and 3. In our previous study (5) 15 individuals out of the group of 20 showed, for both test and retest, a pattern similar to the group pattern shown in the present Table 3; the remaining 5 did not. In the present study 12 out of 20, taken as individuals, showed the same group pattern on both test and retest. The remaining 8 individuals included 6 who showed the group pattern on one occasion but not on the other; and 2 individuals who showed just the opposite pattern, i.e., a decrease in average silence in Period 2 rather than an increase. These 2 "deviates" manifested the same pattern on both test and retest and thus are comparable to the 5 "deviates" of our previous study (5) who likewise were stable in their deviant pattern on the two occasions.

The analysis by individuals as well as the group trends shown in Table 3 are thus a cross validation of our earlier finding that interaction patterns of patients, while invariant under identical inter-interviewer conditions across the interview as a whole (Table 2), are markedly susceptible to planned changes in the stimulus conditions (presented as intra-interview changes from Periods 1 to 5 in the behavior of a single interviewer). Thus Table 3 shows that, as a given interviewer changes his behavior from one period to another, there are predictable associated changes in interviewee interaction patterns. However, Table 2 (as well as Table 3) shows that the same interviewer on retest can elicit almost identical interviewee patterns (invariance) during this interview taken as a whole providing that he presents roughly identical stimulus (or interview) conditions.

In Table 4 can be seen the extent to which the mean patient interaction values shown in Table 3 are reliable from initial test to retest for any one period, in contrast to the interview as a whole. Both rho and r correlation values are shown, although it was demonstrated in our earlier study (5) that the rank-order correlation, rho, is the more appropriate statistic for our data in view of the concomitant problems of extremely narrow ranges, limited frequencies, and some occasional skewing. Values for r are thus included in Table 4 only for purposes of comparison.

In the previous study in which rho was also used, 38 of the total of 45 rhos (9 variables times 5 periods) were significantly reliable; 34 at the .01 level and 4 at the .05 level (5). In

[3] The test-retest raw scores for all the interaction chronograph variables of each of the 20 subjects studied here, as well as a description of the method used to obtain them from the printed interaction chronograph record, can be obtained by ordering Document No. 5095 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $2.00 for microfilm or $3.75 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 4
Test-Retest Correlations for Each of the Subperiods of the Interview

Period	Pt.'s Units		Pt.'s Action		Pt.'s Silence		Pt.'s Tempo	
	rho	r	rho	r	rho	r	rho	r
I	.464*	.490*	.450*	.030	.627**	.532*	.368	.023
II	.428*	.052	.653**	.686**	.520*	.592**	.687**	.688**
III	.669**	.665**	.476*	.153	.465*	.331	.487*	.156
IV	.167	.210	.183	.250	.327	.194	−.060	−.186
V	.371	.439	.280	.265	.514*	.500*	.220	.272
Total	.765**	.890**	.688**	.464*	.642**	.648**	.735**	.764**

Period	Pt.'s Activity		Pt.'s Adjust.		Dr.'s Adjust.		Pt.'s Synchron.		Dr.'s Units	
	rho	r	rho	r	rho	r	rho	r	rho	r
I	.518*	.037	.647**	.202	.536*	.550*	.550**	.249	.336	.458*
II	.553**	.679**	.458*	.572**	.535*	.565**	.376	.208	.584**	.580**
III	.429*	.154	.582**	.398	.458*	.293	.617**	−.062	.572**	.670**
IV	.314	.205	.443*	.439	.211	.123	.613**	.240	.323	.447*
V	.235	.248	.588**	.428	.548**	.507*	.671**	.018	.368	.470*
Total	.738**	.768**	.776**	.757**	.558**	.554*	.544**	.526*	.528*	.823**

* Significant at the .05 level of probability.
** Significant at the .01 level of probability.

TABLE 5
Means, Standard Deviations, Ranges and Rhos for Initiative (Period 2) and Dominance (Period 4)

	# 1	# 2	rho
1. Mean Initiative (Period 2)	.70	.67	.416
SD	.19	.16	
Range	.29 to .92	.38 to .92	
2. Mean Dominance (Period 4)	−.46	−.36	.404
SD	.26	.35	
Range	0 to −.92	+.75 to −.80	

Table 4 of the present study, 29 of the possible 45 period-variable combinations show statistically significant (stable) values of rho; 15 at the .01 level and 14 at the .05 level. Thus, in the subperiods, as well as in the interview as a whole (shown in Table 2), the use of a test-retest interval of seven days, in contrast to several minutes, is associated with somewhat less stability, relatively speaking, for the interaction variables studied.

However, when one remembers the artifacts in measurement described earlier, and also the fact that some of the periods shown in Table 4 are based on only a small fraction of the total (32-minute) interview, the 29 statistically significant rhos out of a possible 45 still indicate considerable stability for the interaction patterns of the 20 patients studied. Thus, for example, the finding of so few statistically significant correlations for Period 4 is probably a function of the fact that this period represents only 2.37 minutes of the total 32-minute interview; the well-known Spearman-Brown phenomenon. Of all the 5 periods shown in Table 4, Periods 3 and 2 elicit, relatively speaking, the most stable values of all nine variables. The stability of Period 3 in this study is in contrast to our earlier finding that this period was one of the least stable (6). However, in our earlier study we had some evidence that this finding was a carry-over of the differences in the way our two interviewers applied the silence stress in Period 2. Since, in the present study, the same interviewer applied the silence stress in both the test and retest of Period 2 this difference in the two experiments is not surprising.

The results of Tables 3 and 4, while somewhat less striking than the comparable findings of our two earlier studies, nevertheless indicate statistically significant (i.e., predictable) modifiability in patient interaction patterns as a function of planned changes in the stimulus conditions presented by the interviewer from one subperiod of the interview to another.

Finally, Table 5 shows the means, standard deviations, ranges, and stability coefficients

for the *Initiative* (Period 2) and *Dominance* (Period 4) variables in the first and second interviews.

The values of rho for both these variables are significant at the .06 level of confidence. These values indicate a fair stability for these interaction variables, in view of the fact that the initiative score is based on Period 2 which was of only about 7.50 minutes' average duration, and the dominance score is based on Period 4 which was of only about 2.37 minutes' average duration. Pearson *rs* for these two variables, .335 and .367, respectively, are slightly lower than the corresponding rhos. Our earlier suggestion (5) would still seem to be pertinent, that the degree of invariance of a single individual's "initiative" or "ascendance-submission" pattern could be investigated in greater detail (and presumably with a corresponding increase in stability) by extending, in a separate study, the length of each of these two periods to perhaps 15 minutes.

Despite the fair stability (rho) for these two variables, the general lowering of the stability coefficients in the present experiment holds also for both of them since three out of four values for the two earlier studies reached the .01 level and the remaining one (*Dominance* in the original study) reached the .05 level. The .06 level of confidence, found for each of the two variables in the present study, is therefore another indication of a somewhat lowered stability coefficient when the test-retest interval is increased from a few minutes to seven days, even if the same interviewer is used for retest rather than a different interviewer.

DISCUSSION

These results in conjunction with those reported in our earlier studies, seem to indicate two things: (*a*) a test-retest interval of several minutes and the use of two different interviewers (5, 6), as well as retest interval of seven days and the use of the same interviewer on the two occasions, both result in the finding of considerable stability for the rather complex interaction variables obtained during a psychologically meaningful, miniature interpersonal situation (the interview). It is, however, apparent that (*b*) the stability is somewhat lower for the second of the two methods employed.

There are undoubtedly many factors associated with this slight lowering of stability in the seven-day study. Four factors, at least, may be involved: (*a*) an habituation effect as any given patient "gets used to" the same interviewer on the two occasions; (*b*) the possibility that the patient's "set" may be different in still another way on the two occasions, e.g., as manifested by a difference in what he talked about (differences in the patient's "content" in the two interviews); (*c*) the possibility of extra-interview influences in the life of the patient during the seven-day interval, e.g., major life changes such as a decision to leave one's spouse, change jobs, etc.; and (*d*) a true relative lowering of stability in time simply as a function of "random" or "unknown" life factors, such that one would expect progressively lower stability coefficients as the retest interval was increased to one month, 6 months, one year, etc.

Each of these four possibilities seems amenable to experimental investigation. The "habituation" factor could be explored by employing, for test-retest, as in our two earlier studies, two different interviewers with the same patient (in counterbalanced order) for interviews seven days apart. If stability coefficients higher than those of the present study were obtained the "habituation" hypothesis would be tenable. Our group is currently investigating the hypothesis that the "content" of the two interviews may be more dissimilar in interviews seven days apart than in two interviews conducted on the same day. The third possibility, that important "extra-interview influences" may modify behavior during the seven day test-retest interval, could be investigated by using any of a number of devices (a specially designed questionnaire, an open-ended interview, etc.) at the end of the retest session. The fourth possibility, that of progressive "decay" in stability coefficients as a function of time, could be investigated by using, in counterbalanced order, two interviewers two weeks apart in one study, one month apart in another, 6 months apart in a third, etc. Of course, if the "habituation" factor were also of interest here, one still could employ the latter approach but use only one interviewer for both test and retest.

In contrast to the four factors just discussed, an alternative hypothesis may explain the slight lowering of stability coefficients in the present study. The standard deviations of all nine of the interaction variables shown in Table 2 for the first interview are smaller than

their counterparts in the two earlier studies. (To take one example, while the average number of patient units in all three investigations is approximately 72, the standard deviation of 15 for the present study is considerably lower than the value of 25 found in the two other series.) The smaller standard deviations indicate that the present sample of patients is more homogeneous with respect to all the interaction variables studied. The lowering of stability coefficients from initial test to retest thus may be merely a reflection of the narrower range of variability for each of the nine variables. Choice among the various possibilities mentioned must await further research.

SUMMARY

The present investigation, an extension of two earlier ones, was a study of the interaction patterns of 20 patients who were interviewed by a single examiner in test-retest interviews seven days apart. The results indicate both that patient interaction patterns are remarkably stable, or invariant, over this seven-day interval, and that the interaction patterns can be predictably modified (within either test or retest interviews) by planned changes in the interviewer's behavior.

REFERENCES

1. CHAPPLE, E. D. The interaction chronograph; its evolution and present application. *Personnel*, 1949, **25**, 295–307.
2. CHAPPLE, E. D. The standard experimental (stress) interview as used in interaction chronograph investigations. *Hum. Organ.* 1953, **12** (2), 23–32.
3. MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1955.
4. MATARAZZO, J. D., SASLOW, G., & MATARAZZO, RUTH G. The interaction chronograph as an instrument for objective measurement of interaction patterns during interviews. *J. Psychol.*, 1956, **41**, 347–367.
5. MATARAZZO, J. D., SASLOW, G., & GUZE, S. B. Stability of interaction patterns during interviews: a replication. *J. consult. Psychol.*, 1956, **20**, 267–274.
6. SASLOW, G., MATARAZZO, J. D., & GUZE, S. B. The stability of interaction chronograph patterns in psychiatric interviews. *J. consult. Psychol.*, 1955, **19**, 417–430.
7. SASLOW, G., GOODRICH, D. W., & STEIN, M. Study of therapist behavior in diagnostic interviews by means of the interaction chronograph. *J. clin. Psychol.*, 1956, **12**, 133–139.

Received May 18, 1956

AUTHORITARIAN ATTITUDES AND PERSONALITY MALADJUSTMENT

ARTHUR R. JENSEN[1]

Institute of Psychiatry (Maudsley Hospital), University of London

SINCE the publication of *The Authoritarian Personality* (2), the relationship of ethnic prejudice and authoritarian attitudes to personality and psychopathology has been the subject of a number of studies. Several scales have been used as measures of authoritarian attitudes, and the personality factors underlying authoritarianism have been investigated with an even greater variety of techniques. The most frequently used attitude measures in studies of this type have been the Anti-Semitism (A-S), Ethnocentrism (E), and Fascism (F) scales developed in *The Authoritarian Personality*. The personality side has been assessed by means of clinical interviews (2), psychoanalysis (1), the Thematic Apperception Test (2), the Rorschach (15, 19), the Rosenzweig P-F test, the Allport-Vernon Scale of Values, and other tests (10). The MMPI has been used most often in studying the authoritarian syndrome (2, 3, 8, 10, 11, 12, 17). A survey of these studies, the majority of which have been reviewed by Christie (5), leaves no doubt that personality factors account for a substantial portion of the variance in ethnic prejudice and authoritarian attitudes in general. The question that still remains open, however, concerns the characteristics and the extent of these personality factors.

Two rather different approaches to this question may be discerned. In one approach the emphasis is on describing the psychodynamics (if the investigator is psychoanalytically inclined) or the trait structure (if the investigator is factor-analytically inclined) of the authoritarian personality, but without reference to psychological maladjustment per se. The dynamic approach is exemplified, for example, by the work of Frenkel-Brunswik (see 5, pp. 226—275). The trait structure approach is represented in the work of Eysenck (7), who conceives of the authoritarian personality as representing one extreme of the dimension of "tough-mindedness-tendermindedness," which in turn he conceives as a projection onto the social attitude field of a set of constitutionally determined personality variables, viz., introversion-extraversion. In Eysenck's scheme the introversion–extraversion dimension is represented as independent of "neuroticism" or maladjustment. More will be said concerning this theory in the light of the present evidence.

The second, more prevalent, approach has been concerned with the kind and degree of personality maladjustment associated with authoritarian attitudes. Psychiatric diagnostic techniques, particularly the MMPI, have figured prominently in this research. And it is largely at this point that most of the disagreements have arisen. Thus Masling (17) has criticised the idea that authoritarianism is related to psychological maladjustment, and has mentioned investigations on four groups of psychiatric patients that failed to show significant correlation between various measures of authoritarianism (Anti-Semitism, Ethnocentrism, and Fascism scales) and several criteria of psychopathology, including the MMPI. And Maria Levinson (2, p. 968) found that the entire range on the Ethnocentrism scale was represented in a group of psychiatric clinic patients.

The studies which have found more positive evidence of a relationship between maladjustment and authoritarianism (8, 10, 11) have been addressed not to the question of whether psychiatric patients obtain higher or lower scores than "normal" subjects on measures of authoritarianism, but rather to that of whether persons in the normal or nonpsychiatric population who score high on measures of authoritarianism show a greater degree of mental ill-health than persons scoring low on authoritarianism. It seems likely that the relationship between mental ill-health and authoritarianism does not hold in both directions. In other words while it may be possible to have any degree of mental illness without showing authoritarian attitudes, it may not be possible to manifest an extreme degree of authoritarianism without being psychologically maladjusted. One may conceive of the scatter diagram of the correlation between authoritarianism and maladjustment as being more the

[1] Public Health Service Research Fellow of the National Institute of Mental Health.

shape of a right triangle rather than the usual ellipse. This rough hypothesis helps to explain some of the seeming contradictions found in the literature, and if it appears at all tenable, as it does to the present writer in viewing the recent research, it would seem that the most satisfactory approach to the problem of authoritarianism and mental health would be to assess the psychological adjustment of individuals at the two extremes on measures of authoritarian attitudes. This essentially was the plan of the present study.

The Prejudice scale. The principal instrument used as a measure of prejudice in this study was the *Pr* (prejudice) scale, which Gough (11) derived by an item analysis of the MMPI, using the Levinson-Sanford Anti-Semitism scale (16) as the criterion. While the *Pr* scale consists of 32 items of the MMPI, it has no more than seven items in common with any other MMPI scale (*Pa*), and one-fourth of the *Pr* items do not appear in any of the other clinical scales.

The reliability and validity of the *Pr* scale were sufficiently high in Gough's studies (11, 12) to warrant its use as an instrument for measuring prejudice on the personality level. The *Pr* scale has been shown to be significantly correlated with anti-Semitic, ethnocentric, and authoritarian attitudes as measured by the A-S, E, and F scales (11). The relative magnitudes of these correlations indicate that the *Pr* scale may be regarded actually as much a measure of authoritarian attitudes as of prejudice. It should also be recognized that the 32 items of the *Pr* scale, constituting about 6 per cent of the total MMPI items, are almost certainly not the only items of the MMPI that are correlated with authoritarianism. Altus and Tafejian (3) subjected the MMPI to an item analysis using high and low scores on the F scale as the criterion, and found 40 items which together correlated .62 with the F scale. Yet only six of the *Pr* items were among these 40. In the present study an item analysis of the *Pr* scale, using *Ss* in the highest and lowest 27 per cent on the F scale as the criterion, showed that 20 of the 32 *Pr* items discriminated between the high and low groups in the F scale at the 5 per cent level of confidence or better. Thus there are at least 66 items of the MMPI which have been shown to be correlated with authoritarian attitudes. It appears, however, that while many of the MMPI items are related to the authoritarian syndrome, the relationship is tenuous and various items do not have the same degree of discriminatory power in different populations. A likely hypothesis seems to be that the discriminatory power of the items is related to the degree of educational sophistication of the subjects and probably other factors that have educationally selective effects. This possibility is suggested by the fact that 20 of the *Pr* items discriminated between high and low F-scale subjects among the San Diego State College students and only 6 of the *Pr* items were found to discriminate between high and low F-scale subjects in the study by Altus and Tafejian (3), whose subjects were students in a psychology course in the Santa Barbara College of the University of California, and probably represent an educationally, or at least psychologically, more sophisticated group. (Psychology majors in San Diego State College generally had the lowest *Pr* scores of any students in the present study.)

The content of the *Pr* items may be interpreted in general terms as representing attitudes of anti-intellectuality ("I like science"—*false*. "I like poetry"—*false*), cynicism ("Most people make friends because friends are likely to be useful to them"—*true*), distrust ("I think most people would lie to get ahead"—*true*), doubt and suspicion ("I commonly wonder what hidden reason another person may have for doing something nice for me"—*true*), misanthropy and querulousness ("The man who provides temptation by leaving valuable property unprotected is about as much to blame for its theft as the one who steals it"—*true*), discontent with self-status and fearful self-concern ("It makes me feel like a failure when I hear of the success of someone I know well"—*true*. "Several times a week I feel as if something dreadful is about to happen"—*true*).

The fact that certain MMPI items differentiate between high and low scorers on measures of prejudice and authoritarianism and that the *Pr* scale is correlated with the A-S, E, and F scales has led some writers (3, 11) to characterize the authoritarian personality in terms of the content of these MMPI items. Strictly speaking this is not a valid prodecure. In order to be valid, an additional investigation would be required to determine whether or not the items were factually true statements about the individuals taking the test. Since this has

never been done, and indeed would not even be advisable, we can legitimately make no use of item content other than as a possible source of hypotheses about personality.

A more defensible approach would consist of determining the relationships between the authoritarian measures and the various clinical scales of the MMPI and then characterizing the authoritarian personality in terms of our empirical knowledge of the meaning of the clinical scales, which have been elucidated by innumerable studies (6). Even then, caution must be taken not to think of the authoritarian personality in the singular. The approach suggested above would permit a number of generalizations about the personality characteristics to be found among a *group* of persons having authoritarian attitudes, but most likely only some of these generalizations would hold true for any individual.

METHOD

Subjects. The group form of the MMPI was administered to 712 entering freshmen in the San Diego State College, a four-year coeducational liberal arts college in Southern California, and was readministered one year later to the 312 students of this class who remained in college to begin their sophomore year. The same data and in addition the California F scale (Form 60A) were obtained from 114 seniors majoring in education. Intelligence test (American Council on Education) scores and grade-point averages were also available for all the freshmen.[2]

The one rather atypical feature in this sample seems to be the high drop-out rate (56 per cent) between the freshman and sophomore years. The writer can only add his speculations to the few points in the data that suggest possible causes for so many drop-outs. For one thing, the "G.I. Bill" was still in effect at the time these data were gathered and large numbers of men were caused to enter college who otherwise would not have sought college training, not only for lack of financial means but also of the necessary academic interests and aptitudes. The college entrance requirements were relaxed for these students, while academic standards were maintained at a high level. Thus many were discouraged from continuing beyond their first year. The mean *Pr* score of the students who dropped out of college was significantly ($p < .05$) higher than those who continued, suggesting that the drop-outs were more like noncollege groups and high school students who do not plan to go to college (13, p. 265). These groups typically are known to obtain higher *Pr* scores on the average than college groups.[3] Another cause of the high drop-out rate is that for many students the state college is a stepping stone to other colleges and universities.

[2] I am indebted to Dr. Kenneth Eells for having made these data available to me.

[3] The writer also administered the *Pr* and F scales to 96 junior college vocational students in the same locality and found their mean scores on both scales to be significantly ($p < .001$) higher than those in the liberal arts college.

TABLE 1

T CONVERSION TABLE AND PERCENTILES FOR *Pr* SCORES, BASED ON 712 COLLEGE FRESHMEN (342 MALES, 370 FEMALES)

Raw Score	*f*	Percentile	*T* Score*	Raw Score	*f*	Percentile	*T* Score
28			**	13	23	83	60
27	1	99	77	12	43	79	58
26			76	11	40	73	56
25			76	10	50	67	55
24			75	9	41	60	53
23			75	8	66	53	51
22	1	99	74	7	62	44	48
21	7	99	73	6	64	35	46
20	2	98	71	5	55	27	44
19	2	98	70	4	56	19	41
18	13	97	69	3	48	12	38
17	14	95	67	2	41	5	34
16	14	93	65	1	11	2	29
15	24	90	63	0	7	1	24
14	27	87	61				

* Since the distribution of raw scores is significantly skewed, the *T* scores have been normalized.

** The absence of frequencies above this point prohibits the computation of *T* scores.

RESULTS

Characteristics of Pr. Since normative data on the *Pr* scale have been published only for a sample of high school seniors (11, 12), it would seem advisable to present similar data for a college sample before going on to make further interpretations on the basis of *Pr*.

The distribution of raw *Pr* scores in the college sample is conspicuously more constricted and positively skewed than in Gough's high school sample. In order to make comparisons of the *Pr* scale with other MMPI scales, a *T* conversion table was prepared (Table 1). This *T* scale has a mean of 50 and a standard deviation of 10. Since a test of the skewness of the distribution of raw scores was significant at the 5 per cent level, the *T* scale was normalized.

Table 2 is quite self-explanatory. The following points in this table are to be noted.

a. The reliability of *Pr* is consistent with Gough's findings (11, p. 254) and compares favorably with the reliability of other MMPI scales (see 6, p. 5).

b. The retest reliability after one year is notably high considering the restricted range of scores.

c. The mean *Pr* for the freshmen is approximately one standard deviation lower than that

TABLE 2
DATA ON *Pr* SCALE IN TWO COLLEGE SAMPLES

Pr Data	College Freshmen $N = 712$	Education Majors (Seniors) $N = 114$
a Split-half reliability (odd-even, corrected by Spearman-Brown formula)	.81 ($N = 100$)	
b Retest reliability (1 year interval)	.56 ($N = 312$)	
c Mean and Standard Deviation	8.35 *SD* 4.59	6.37 *SD* 3.43
d Average decrease in *Pr* after 1 year of college and Critical Ratio for the difference	2.05 *CR* 9.45 ($N = 312$)	
e Correlation of *Pr* with *F* scale (Form 60A)*		.27 SE .09
f Correlation of *Pr* with intelligence (American Council on Education test)	−.15, SE .06 ($N = 288$)	
g Correlation of *Pr* with Grade Point Average	−.13, SE .06 ($N = 288$)	

* This scale is identical with the E-F scale given in Gough paper (10, pp. 239-241).

of Gough's high school sample, and the mean of the education majors is nearly two standard deviations lower. This is consistent with the findings of other studies, which report significant correlation between educational level and measures of authoritarianism (5, p. 170).

d. The systematic lowering of *Pr* scores after one year of college is highly significant. While 69 per cent of the *Ss* obtained lower *Pr* scores on the retest, only 20 per cent obtained higher scores. The 11 per cent whose scores did not change over the one-year interval had a very low mean *Pr* of 5.77, *SD* 3.71.

e. The correlation of *Pr* with the F scale in the college senior group is considerably lower than the correlation of .46 in Gough's study. This is undoubtedly due to the very restricted range of scores on both scales in the college sample, a factor tending to lower the correlation coefficient. When the present sample was combined with a group of 96 junior college vocational students who had taken the same tests, the correlation between *Pr* and F rose to .65, SE .07.

f & *g*. The correlations of *Pr* with intelligence and grade-point average are similarly much lower than those in Gough's study, again because of the greater homogeneity of the college sample, with the consequent restriction of range on the correlated variables. Both the correlations, however, are statistically significant at better than the 5 per cent level of confidence. While there is certainly a negative correlation between intelligence and *Pr*, as well as of other measures of authoritarianism (5, p. 168), intelligence may be regarded as a negligible factor in the variance of *Pr* in a college population. For example, it was found that groups even with widely differing *Pr* scores (above 15 and below 3) showed no significant difference in mean intelligence test (ACE) scores.

Some additional findings about *Pr* may be noted. There were no significant differences between men and women. Students majoring in different subjects had significantly different mean *Pr* scores. Why these differences should exist among groups of students of the same age, at the same stage in their education, and in the same college would be difficult to explain except in terms of personality differences. Most of the groups representing different fields of study were too small to permit meaningful statistical comparisons, but the two largest groups—business majors and education majors—showed a highly significant difference ($p < .001$) in mean *Pr* scores. The high *Pr* groups were students majoring in business administration, commerce, social work, art, music, mathematics, and chemistry. The low *Pr* groups were students majoring in psychology, education, physics, biology, and economics. Any generalizations or interpretations from these data alone would necessarily be *ad hoc*. Since the students in all these groups took the MMPI at the very beginning of their freshman year, there can be no reflection on the influence of the faculties or curricula of the various departments of the college in contributing to these differences in mean *Pr*.

Pr *and personality adjustment*. The greatly restricted range of the *Pr* distribution in this college sample, with the large piling-up of scores toward the low end of the scale, and the likely constriction and skewness of the distribution on most of the clinical scales of the MMPI, deterred the investigator from studying the relationship between *Pr* and the clinical scales by means of the correlation co-

efficient. The nature of the distributions in this sample would have so attenuated the correlations as to obscure possibly important relationships. Therefore it was decided to compare the MMPI clinical scales of adequately large groups of high and low *Ss* on *Pr* to permit the emergence of relatively unambiguous statistical significance should real differences exist. While it is the standard practice to use the upper and lower 27 per cent of the distribution in comparing high and low groups, this cutoff point would be impossible in the present distribution, as is obvious from inspection of the Percentile column in Table 1. There would be only five points difference between the highs and lows if the upper and lower 27 per cent were used, and furthermore the highs would still fall below the mean *Pr* of noncollege groups. Therefore it was decided to select from the 712 freshmen who took the MMPI the upper 7 per cent ($N = 54$) on the *Pr* scale (scores above 15) and the lower 8 per cent (scores below 3, $N = 59$).

Comparisons between these groups on the nine clinical scales and the *K* scale of the MMPI revealed that some of the scales discriminated positively, some negatively, and others not at all. On each scale that showed a difference between the high and low *Pr* groups, this difference was significant at better than the .001 level of confidence. The differences on the remaining scales were in all cases smaller than the standard error of the difference and thus completely nonsignificant. The high *Pr* group was higher (more "maladjusted") on the following MMPI scales: *Hs* (hypochondriasis), *D* (depression), *Pd* (psychopathic deviate), *Pt* (psychasthenia), *Sc* (Schizophrenia), and *Ma* (hypomania). The following scales discriminated negatively: *K* ("defensiveness") and *Hy* (hysteria). The *Mf* (masculinity–femininity) and *Pa* (paranoia) scales did not discriminate significantly. These results are presented in Table 3 in such a way as to permit comparisons with other studies on the relationship between measures of authoritarian attitudes and the MMPI scales. The magnitude of the correlations has not been given, as

TABLE 3

RELATIONSHIPS[1] BETWEEN VARIOUS MEASURES OF AUTHORITARIANISM AND THE MMPI SCALES

Study	Sample	*N*	Measure of Relationship	Criterion Scale	MMPI Scales										
					K	*Hs*	*D*	*Hy*	*Pd*	*Mf* (m)	*Mf*(f)	*Pa*	*Pt*	*Sc*	*Ma*
Jensen	College Freshmen	113	β	*Pr*	−	+	+	−	+				+	+	+
Tyler (20, p. 453)	Graduate Education Majors (female)	107	r	*Pr*	0[2]		+	−				−	+	+	+
Gough (12, p. 258)	High School Seniors[3] Class A	271	r	*Pr*	−		+	−	+	+		+	+	+	+
Gough (12, p. 258)	High School Seniors Class B	231	r	*Pr*	−		+		+		+		+	+	+
Gough (10, p. 241)	High School Seniors Class A[4]	271	r	A-S[4]	−		+			−	−			+	+
Gough (11, p. 249)	High School Seniors Class A	80	t[5]	A-S[4]	−		+		+					+	
Gough (10, p. 241)	High School Seniors Class A	271	r	E-F[6]	−			−		−					
Freedman, *et al.* (8, p. 317)	College Freshmen Sample 1 (female)	441	r	F[7]	−	+		−[8]				−	+		
Freedman, *et al.* (8, p. 317)	College Freshmen Sample 2 (female)	225	r	F[7]	−			−[8]				−			−

[1] + indicates a positive correlation, − indicates a negative correlation. Only relationships significant beyond the .05 level of confidence are given. In the Jensen study all relationships are significant beyond the .001 level.

[2] t test for significance of differences between means of highest 7 per cent (*Pr* scores above 15) and lowest 8 per cent (*Pr* scores below 3) in a sample of 712 *Ss*.

[3] *K* scale not used in Tyler's study.

[4] Levinson-Sanford Anti-Semitism Scale (16).

[5] t test for significance of differences between means of highest 40 and lowest 40 on A-S scale in a sample of 271 *Ss*.

[6] Gough's *E-F* scale (10, pp. 239-241) is made up of 30 items from the California F scale, Form 60 (2, pp. 248-250).

[7] California F scale (2, pp. 255-257).

[8] The somatic items of the *Hy* scale were not counted for this correlation in Freedman's study. The correlation for the complete *Hy* scale was non-significant.

[9] The Class A High School Seniors in Gough's study is the same sample in every instance; therefore the four sets of data from Gough are not independent.

these vary considerably from one sample to another and essentially add nothing to the picture. Only those relationships significant beyond the 5 per cent level have been included.

Though these results clearly establish the fact of a relationship between *Pr* and certain clinical scales, they alone are not sufficient to answer the question concerning the relation of *Pr* to psychological maladjustment. The correlation, or differences between high and low groups, could conceivably be based entirely on MMPI scores considered to be well within the normal range of psychological adjustment. In order to answer this question the high and low *Pr* groups were compared with respect to the proportion of each group obtaining *T* scores above 70, i.e., the level generally indicative of psychological maladjustment. The results were clear-cut. A significantly ($p < .01$) greater proportion of the high *Pr* group obtained *T* scores above 70 on the following scales: *D*, *Pd*, *Pa*, *Pt*, *Sc*, *Ma*. For each of these scales, on the average, there were about twelve times as many of the high *Pr* group who obtained *T* scores above 70 as of the low *Pr* group. The scales on which the proportions of high and low *Pr* groups having *T* scores above 70 were *not* significantly different were *Hs*, *Hy*, and *Mf*.[4] It should be noted that while in the comparisons of Table 3 the *Hy* scale is inversely related to *Pr*, the high and low *Pr* groups do not differ in the proportions obtaining *T* scores above 70 on the *Hy* scale. This finding suggests that though low *Pr* is associated with high *Hy* in this sample, the *Hy* is not so high as to be in the abnormal range. The nonhomogeneous character of the *Hy* scale also enters the picture here. For example, Freedman *et al.* (8) found a significant negative correlation between *Hy* and the California F scale in their college sample only when they removed the somatic items from the *Hy* scale.

Another question remains to be answered concerning the relationship of *Pr* to adjustment. That is, do students who show signs of poor psychological adjustment as judged independently of the MMPI also have higher *Pr* scores than students who appear to be well-adjusted? To get at least a rough answer to this question the entire college faculty was asked to submit the names of students whom they knew personally and considered either well-adjusted or poorly adjusted in the sense of being likely candidates for psychological counseling. None of the faculty had any knowledge of the aims of this study. Students about whom there was general agreement by two or more of the faculty were compared on the *Pr* scale. Again the results were clear-cut. The poorly adjusted group ($N = 30$) had a significantly ($p < .001$) higher mean *Pr* (9.93) than the well-adjusted group ($N = 57$) with a mean *Pr* of 5.61.

[4] The *K* scale was not included in these comparisons since it has not been validated as a measure of maladjustment.

DISCUSSION

There is quite clearly a relationship between authoritarian attitudes on the personality level as measured by the *Pr* scale and psychological maladjustment as measured by the MMPI as well as by faculty ratings. But in addition to this general finding, what more specifically can be gleaned from these data as to the aspects of personality tapped by the *Pr* scale? While in any one study in which a number of correlations or group differences are obtained certain of these may lack general significance because of peculiarities of the particular sample, the type of analysis applied, or the merely chance fluctuations of sampling, the marked consistencies shown in the studies summarized in Table 3 do provide a sound basis for interpretation. The *K*, *D*, *Hy*, *Pt*, and *Sc* scales display the most consistent relationship to measures of authoritarianism, and so the discussion is confined to these.

Not only does *K* show negative correlations with authoritarian measures in every study, but these correlations are generally higher than any of the others. Gough (12), for instance, found correlations between *Pr* and *K* which were nearly as high as the reliabilities of either of these scales. The meaning of *K* is complex. The scale was originally devised as a correction factor for the clinical scales (18) but has since been found to have psychological significance in its own right. In the clinical interpretation of the MMPI a high *K* score is regarded as indicative of a defensive attitude, and a low *K* score suggests unguarded and indiscrete frankness and a tendency to speak critically of self. Low *K* is found most frequently in conditions characterized by ego

weakness, in which the person is psychologically naked and vulnerable, as in the psychoses and anxiety states. *K* shows a high negative correlation (−.74) with the Taylor Manifest Anxiety Scale (4), and a factor analysis of the MMPI in a college sample by Wheeler (21) shows a low *K*, along with high *Pt* and *Sc*, to be highly saturated with the factor interpreted as anxious self-concern with which the ego defenses seem inadequate to cope. High *K*, along with high *Hy*, was saturated with the factor interpreted as intactness of the ego-defense mechanisms. Wheeler's study also highlights the inverse similarity between *K* and *Pr*, in that correlations between *K* and the other MMPI scales show very much the same pattern as does the *Pr* scale, except of course that the direction of the correlation is reversed, since *Pr* and *K* are negatively correlated. High *K* scores have been found to be associated with the following traits: sociable, wide interests, reasonably enthusiastic, and verbal; low *K* scores: high-strung, cynical, dissatisfied, and individualistic (14, p. 78). *K* is positively correlated with intelligence and socioeconomic status and there is a tendency for college students and college-educated persons to obtain *K* scores one-half to one standard deviation higher than the noncollege population—the same relationships that have been found (inversely of course) for the *Pr* scale (6, 15). Of all the MMPI scales, *K* is probably the best single index of the personality factor tapped by *Pr*, at least in the nonpsychiatric population.

The positive correlation between ethnic prejudice and the *D* scale was first noted in *The Authoritarian Personality*, somewhat to the surprise of the investigators, since clinical signs of depression were more apparent in psychiatric patients scoring *low* on ethnocentrism. As seen in Table 3, however, the positive relationship of authoritarian attitudes with *D* is entirely consistent in several studies. That *D* is not a homogeneous scale and must always be interpreted in relation to other scales in clinical practice may account for the apparent discrepancies with the clinical observation noted in *The Authoritarian Personality*. There is considerable overlap in the traits associated with high *D* and low *K*. Persons obtaining high *D* scores are characterized by lack of self-confidence, tendency to worry, narrowness of interests, poor morale, an uneasy self-concern and dissatisfaction with their current situation (6). Except in relation to other scales *D* is difficult to interpret clinically. However, taken alone it may be regarded as the best single index of maladjustment in the MMPI (9).

The *Hy* scale shows a consistently negative correlation with the *Pr* and F scales. Since *Hy* and *K* are highly correlated, a similar interpretation as that given for *K* applies also to *Hy*: in short, there appears to be a negative relationship between repressive tendencies and authoritarianism. Like *K*, *Hy* is positively correlated with intelligence and is usually about one-half standard deviation higher in college samples than in the adult standardization group. It is quite clearly the psychic and not the somatic items of the *Hy* scale that account for its negative correlations with both *Pr* and F. Freedman *et al.* (8) found the correlations between the *Hy* scale and the California F scale in two college samples to be −.02 and .00. But when the somatic items were removed, the correlations rose to −.41 and −.44. A high *Hy* score based on the psychic items indicates a tendency not to tire easily, not to be depressed, a feeling that life is good and that the world is benevolent—in short, a rather optimistic, extraverted attitude.

The negative correlation (or absence of correlation) between authoritarianism and the *Pa* (paranoia) scale is usually viewed with surprise, for *The Authoritarian Personality* called upon the paranoid mechanisms of projection, extrapunitiveness, and power orientation in explaining the authoritarian syndrome. The reason for the negative correlation between *Pa* and F (and presumably also for the lack of correlation with *Pr* in the present study) has been adequately explained by Freedman *et al.* (8) in terms of the heterogeneous nature of the *Pa* items. When the *Pa* scale is broken down into three types of items, the correlations between F and these sub-*Pa* scales are as follows: Persecutory Ideas ($r = .01$), Poignancy ($r = -.07$), and Naïveté ($r = -.22$). This last correlation was significant at the .001 level and it seems safe to say it is this element of the *Pr* scale that largely accounts for the negative correlation with authoritarianism. The "persecutory"

items are not at all subtle, hardly concealing their pathological implications, and comparatively few of them are responded to in the keyed direction by more sophisticated *Ss* such as college students. But as the *Pa* score rises, more of the "persecutory" items are of course included, and it should be noted that when the high and low *Pr* groups in the present study were compared for the proportions of *Ss* obtaining *T* scores over 70, the high *Pr* group was significantly *higher* on *Pa*, because more of them responded in the keyed direction to a larger number of *Pa* items, including the "persecutory" items, which are most commonly associated with paranoia.

The *Pt* scale is a relatively pure measure of neuroticism of the anxiety and obsessive-compulsive varieties. Persons scoring high on *Pt* show excessive doubts, compulsions, obsessions or unreasonable fears, anxiety, perplexity, and apprehension.

The *Sc* scale in nonpsychiatric populations is associated with schizoid tendencies, feelings of isolation, disinterest, and pessimism.

In the over-all picture presented by these scales—those most consistently related to authoritarian measures—the emphasis is on the obsessive-compulsive syndrome (high *D*, *Pt*, *Sc*) and there is comparatively little hysterical and repressive tendency (low *Hy* and *K*). The impression is that prejudiced, authoritarian persons have less well-developed ego defenses and are thus more exposed and vulnerable to psychological stress, in the face of which they develop tendencies toward pessimism, cynicism, low morale (*D*), and psychological isolation (*Sc*), along with the more primitive defenses of a compulsive, ritualistic, and schizoid nature (the triad *D*, *Pt*, *Sc*).

In view of the picture presented by the data of this study, as well as the consistencies seen in several others, it seems to this writer that an adequate theory of social attitudes, particularly those in the domain of ethnic prejudice and authoritarianism, must take into account psychopathological aspects of personality. To ignore them seems an unwarranted oversimplification of this complex field. Eysenck, for example, has formulated a theory of social attitudes which essentially leaves out of account the relationship between social attitudes and personality adjustment (7). Eysenck conceives of social attitudes as being determined by two orthogonal factors—the Radicalism–Conservatism dimension (R-factor) and the Toughminded–Tenderminded (T-factor) continuum. We shall be concerned here only with the T factor, which is regarded as a projection onto the social attitude field of a fundamental dimension of personality, viz., introversion–extraversion which is said to be constitutionally determined and is conceived as being entirely unrelated to emotional instability or neuroticism (7, p. 179). Eysenck equates the authoritarian–democratic continuum with the T factor, authoritarian attitudes being an expression of toughmindedness. Furthermore he considers the A-S, E, F, and *Pr* scales as essentially measures of toughmindedness (7, pp. 148, 152, 233). They should therefore be highly correlated with measures of extraversion and show comparatively little, if any, correlation with measures of neuroticism. Eysenck has also stated that those toward the extraverted end of the continuum develop symptoms of the hysterical type during periods of emotional instability, while those toward the introverted end develop "dysthymic" disorders (anxiety, depression, and obsessive-compulsive).

If the writer understands Eysenck correctly, these relationships postulated by his theory are not in accord with the facts presented here. The measures of authoritarianism were shown to correlate negatively with the *Hy* (hysteria) scale and positively with the "dysthymic" scales (*D*, *Pt*), both correlations being contrary to predictions from Eysenck's theory. Two factor-analytic studies provide similar contradictory evidence as well as demonstrate that measures of authoritarianism have a considerable loading on the neuroticism factor. Tyler (20) performed a centroid factor analysis of fifteen MMPI scales, including *Pr*, on a college sample and found that *Pr* had its most substantial loading (.70) on the first factor, which he interpreted as "general maladjustment." The scale showing the highest correlation with *Pr* was *Pt*, which Eysenck considers to be a relatively pure measure of neuroticism. A Lawley-type factor analysis was carried out in this laboratory by Dr. M. I. Soueif on several MMPI scales (K,

Hy, *Pd*, *Pt*) and Guilford's factorially pure *D* (depression), *C* (cycloid) and *R* (rhathymia scales. (The *R* scale is regarded by Eysenck as one of the best measures of extraversion.) The first two factors in Soueif's analysis were neuroticism and introversion–extraversion. The MMPI scales with the largest factor loadings on neuroticism were *K* (−.64) and *Pt* (.85), both of which are highly correlated with authoritarianism. The *K* and *Pt* scales had comparatively small factor loadings on introversion-extraversion (.40 and .14 respectively). The *Hy* scale had a loading of .12 on the neuroticism factor and .77 on extraversion. But it has consistently shown *negative* correlations with measures of authoritarianism, which is exactly the opposite of what one would predict from Eysenck's theory.

SUMMARY

The MMPI was administered twice, with a one-year interval, to 712 college freshmen and to 114 seniors, who also took the California F (Fascism) scale. A discussion of the nature of the *Pr* (prejudice) scale of the MMPI and its relation to a number of other variables in a college sample are presented, along with a normalized *T* conversion table for *Pr*.

1. The distribution of *Pr* scores in the college sample was more skewed and constricted, with a piling up of low scores, than the distribution of scores in the high school group used in the derivation of the *Pr* scale.

2. The reliability of *Pr* (.81) compares favorably with that of other MMPI scales.

3. There was a significant decrease in mean *Pr* as students advance in college.

4. There were significant differences in mean *Pr* among students majoring in different fields.

5. *Pr* correlated significantly with the California F scale.

6. *Pr* was positively correlated with the *Hs*, *D*, *Pd*, *Pt*, *Sc*, and *Ma* scales, and negatively correlated with *K* and *Hy*.

These findings and those of several related studies are discussed as a means of psychologically characterizing the authoritarian personality. It was concluded that certain psychopathological factors are present in the authoritarian syndrome.

REFERENCES

1. ACKERMAN, N. W., & JAHODA, M. *Anti-Semitism and emotional disorder*. New York: Harper, 1950.
2. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
3. ALTUS, W. D., & TAFEJIAN, T. T. MMPI correlates of the California E-F Scale. *J. soc. Psychol.*, 1953, **38**, 145–149.
4. BRACKBILL, G., & LITTLE, K. B. MMPI correlates of the Taylor scale of manifest anxiety. *J. consult. Psychol.*, 1954, **18**, 433–436.
5. CHRISTIE, R., & JAHODA, MARIE (Eds.). *Studies in the scope and method of "The authoritarian personality."* Glencoe, Ill.: Free Press, 1954.
6. COTTLE, W. C. *The MMPI, A review*. Lawrence, Kansas: Univer. of Kansas Pub., 1953.
7. EYSENCK, H. J. *The psychology of politics*. New York: Praeger, 1954.
8. FREEDMAN, M., WEBSTER, H., & SANFORD, N. A study of authoritarianism and psychopathology. *J. Psychol.*, 1956, **41**, 315–322.
9. GOUGH, H. G. A new dimension of status: II. Relationship of the *St* scale to other variables. *Amer. sociol. Rev.*, 1948, **13**, 534–537.
10. GOUGH, H. G. Studies of social intolerance: I. Some psychological and sociological correlates of anti-Semitism. *J. soc. Psychol.*, 1951, **33**, 237–246.
11. GOUGH, H. G. Studies of social intolerance: II. A personality scale for anti-Semitism. *J. soc. Psychol.*, 1951, **33**, 247–255.
12. GOUGH, H. G. Studies of social intolerance: III. Relationship of the *Pr* scale to other variables. *J. soc. Psychol.*, 1951, **33**, 257–262.
13. GOUGH, H. G. Studies of social intolerance: IV. Related social attitudes. *J. soc. Psychol.*, 1951, **33**, 263–269.
14. HATHAWAY, S. R., & MEEHL, P. E. The Minnesota Multiphasic Personality Inventory. In *Military clinical psychology*, TM 8-242, AFM 160-45. Washington: U. S. Government Printing Office, 1951.
15. HIMMELHOCK, J. The dynamics of tolerance. Unpublished doctor's dissertation, Columbia Univer., 1952.
16. LEVINSON, D. J., & SANFORD, R. N. A scale for the measurement of anti-Semitism. *J. Psychol.*, 1944, **17**, 339–370.
17. MASLING, J. M. How neurotic is the authoritarian? *Amer. Psychologist*, 1953, **8**, 402–403. (Abstract)
18. MCKINLEY, J. C., HATHAWAY, S. R., & MEEHL, P. E. The Minnesota Multiphasic Personality Inventory: VI. The *K* scale. *J. consult. Psychol*, 1948, **12**, 20–31.
19. REICHARD, SUZANNE. Rorschach study of prejudiced personality. *Amer. J. Orthopsychiat.*, 1948, **18**, 280–286.
20. TYLER, F. T. A factorial analysis of fifteen MMPI scales. *J. consult Psychol.*, 1951, **15**, 451–456.
21. WHEELER, W. M., LITTLE, K. B., & LEHNER, G. F. The internal structure of the MMPI. *J. consult. Psychol.*, 1951, **15**, 134–141.

Received May 28, 1956

FACTORS DETERMINING THE GUESSING OF MEANINGS OF FOREIGN WORDS[1]

YVONNE BRACKBILL AND KENNETH B. LITTLE

Stanford University

IN A recent article, Brown, Black, and Horowitz (1) report an experiment in which 86 English-speaking subjects were asked to guess the English meanings of 21 pairs of antonyms presented in three foreign languages—Hindi, Czech, and Chinese. Inter-subject agreement as to meaning of the foreign words ran as high as 97 per cent and, furthermore, the number of correct guesses for each language was significantly greater than chance expectation. The authors interpret their results as supporting the hypothesis that there is some sort of intrinsic correspondence between the sounds of certain words and their meanings regardless of language. They suggest that, "It is possible . . . that the origin of speech in gestural imitation and physiognomic language has left a residue of roughly translatable words in all languages. . . ," and conclude, "For the present we prefer to interpret our results as indicative of a primitive phonetic symbolism deriving from the origin of speech in some kind of imitative or physiognomic linkage of sound and meanings."

The results presented by Brown, Black, and Horowitz are unequivocal in that subjects certainly guessed the correct English equivalents of the foreign words with a surprising degree of accuracy. There are certain aspects of the study, however, that make it difficult to determine whether these results support an explanatory hypothesis of a universal sound-meaning linkage:

1. The method used by the authors was not a fully appropriate test of the hypothesis. Two points may be made in this regard. First, Brown, Black, and Horowitz used visual as well as auditory stimuli in their method of presentation. However, the hypothesis concerns a correspondence between meaning and *sound*, not between meaning and the written form of a word. Therefore, the method most appropriate to a test of the hypothesis would be one depending exclusively on auditory cues. Second, the hypothesis is necessarily concerned with meanings that have some intercultural communality. It does not apply to associational meanings of the type that are largely unshared, i.e., private or idiosyncratic, or that are shared only by a very narrow segment of the population. With English-speaking subjects, the most adequate methodological control for association meanings lies in avoiding the use of English words, using instead a method that pairs one foreign sound against a second foreign sound, with no meaning provided.

2. The authors made no check on the fidelity of their foreign translations. Invariably there are concepts in one language that have no exact and/or single equivalent in another (cf. 4, p. 191). When a translator must choose that particular word which best approximates the intended meaning, it is quite possible that his choice may simply be in error or that he may select that alternative which most closely resembles the English word in length and sound. Hence, it is desirable to have some consensus of translators as well as a back-translation.

3. The population of concepts is indeterminant and the sampling procedures are ambiguous. Brown, Black, and Horowitz state, "English antonyms were selected from the Thorndike-Lorge word list with two considerations in mind. (*a*) The words should name sense experiences (e.g., warm-cool, heavy-light). (*b*) The members of a pair should both fall in the frequency range of 100 or over per million. The final list of 21 pairs consisted of those on which the three authors could agree" (1, p. 389). Yet there are certain discrepancies between their criteria and their actual word sample: (*a*) Thirteen of the forty-two words have frequencies of use less than 100 per million, by the Thorndike-Lorge word list; (*b*) three of the word pairs are clearly not antonyms (thunder-lightning, iron-gold, and

[1] This study was supported in part by a Stanford University Research Grant.

many-one); and (*c*) the sense experience character of the words is quite debatable.[2]

4. Finally, the authors suggest that some universal sound-meaning relation exists in *all* languages. However, those used in their study and in the unpublished studies they cite constitute a small and unrepresentative sample of the population of existing languages. Obviously there are practical limitations to securing a representative language sample. Possibly the best approximation lies in structuring the sample so that it contains not more than one member of any language group. The sample languages, that is, should be noncognate. Unfortunately, three of the four languages used by Brown, Black, and Horowitz (English, Czech, and Hindi) *are* members of the same language group.

Maltzman, Morrisett, and Brooks (3) have reported a series of experiments similar to that of Brown, Black, and Horowitz. In their study, subjects were presented, in printed form only, with a single stimulus word and two response words. The languages used were English, Japanese, and Croatian rotated in position as stimulus and response. The sample of concepts used was that of Tsuru and Fries (5) which is a mixture of antonyms and contradictory word pairs. Their results, in English-Foreign comparisons, were similar to those of Brown, Black, and Horowitz, but their subjects were unable to match Japanese and Croatian words of the same meaning better than chance. They conclude that the latter finding is contrary to the Tsuru and Fries hypothesis of Gestalt organization of trace systems and related hypotheses of phonetic symbolism and physiognomic language such as that advanced by Brown, Black, and Horowitz. As in the Brown, Black, and Horowitz study, the concept population is indefinite and visual stimuli rather than auditory were used.

The hypothesis of a universal sound-meaning linkage based upon a primitive phonetic symbolism or physiognomic language is a nebulous one at best and does not seem to be susceptible to rigorous test in its general form. The results of the two studies cited appear, however, to constitute incontrovertible evidence that naive subjects can guess the English meanings of *certain* words from *some* foreign languages well above chance. The experiments reported below were designed to study aspects of the process of matching words of similar meanings from different languages. Specifically, we attempted to determine: (*a*) the effects on accuracy of guessing of mode of presentation (auditory, visual, or both), (*b*) whether the previously reported above-chance accuracy of guessing would occur when the word sample was a random one from a determinate population, and (*c*) the specific stimulus characteristics that lead to accurate guesses.

[2] As a check, the 42 words in the Brown, Black, and Horowitz list were presented to 28 psychologists with instructions to check those words which they believed named sense experiences. On only 16 of the words did 75 per cent or more of the subjects agree that the term named a sense experience.

METHOD

Four major noncognate languages were used: Hebrew, Japanese, Chinese, and English. Choice of the first three was dictated largely by the availability of several skilled foreign-born translators for each of these languages.

The experimental population of words consisted of concepts of high frequency of occurrence. Ideally, a semantic frequency list of words for each language studied would form the basis for sample selection. However, since no such lists were available for the three foreign languages used, the closest approximation was to employ a list of the most frequently used concepts in four Indo-European languages: English, Spanish, French, and German (2). For the preliminary sample, those words were selected which were (*a*) principal parts of speech—either noun, pronoun, adjective, verb, or adverb, and (*b*) among the 500 most frequently used words in at least three out of the four languages. The 324 words obtained were typed and given to nine translators (three translators per foreign language). The translators were foreign-born, but were thoroughly familiar with the English language. They were instructed to work independently, to use the first person singular in translating verbs, to select the most commonly used expression when alternatives existed, and to use the colloquial rather than literary form. As a check on the appropriateness of the sample source, they were also asked to indicate those concepts which were uncommon in their own tongue.

On completion of each set of translations, the experimenters met with the three translators and compiled a composite list. Eliminated were those words marked as uncommon and those words upon which the translators could not easily reach agreement. The remaining terms were printed in romanized script by one of the translators and submitted by the experimenters to a fourth translator who back-translated them into English. In this way homonyms as well as other words which were not correctly back-translated were also eliminated. A final master list of words was drawn up

TABLE 1
VARIANCE TABLE FOR THE MATCHING OF ENGLISH AND FOREIGN WORDS

Source	*df*	Sum of Squares	*F*	*p*
Individuals	117	1164.6		
Languages	2	167.9	7.46	$<.001$
Method of presentation	2	91.8	4.62	$<.05$
Languages × methods	4	72.9	1.62	$>.10$
Remainder	234	2636.6		
Total	359	4133.8		

by coordinating the separate lists for Japanese, Chinese, and Hebrew. From this alphabetized master list a sample of 50 words was taken by selecting every *n*th word.

A pilot study using an arbitrary list of 36 pairs of antonyms indicated that subjects did as well at guessing meanings when each English word was presented individually with either its foreign equivalent *or* opposite, as when the English and foreign antonyms were presented in pairs. In the interests of economy of both time and stimuli the simpler form was used in this experiment.

Three methods of presentation were used: auditory only (A), auditory and visual (AV), and visual only (V). Auditory presentation was by tape recording in which an announcer spoke the number of the word-pair and two native speakers the stimulus words. Approximately one second elapsed between presentation of the first and second words. An interval of five to seven seconds was allowed for the subject to make his decision. This interval had been found to be more than adequate in the pilot study mentioned above. For visual presentation, foreign words were typed in romanized form.

Each of the three methods of presentation was employed under two conditions. In one (E-F) the word pairs were composed of one English word and one foreign word: English-Japanese (E-J), English-Chinese (E-C), and English-Hebrew (E-H). In the other condition (F-F), one word came from one foreign language and the other word from a different foreign language: Chinese-Japanese (C-J), Chinese-Hebrew (C-H), and Hebrew-Japanese (H-J).

Six separate groups of subjects were used, corresponding to the six method-condition combinations. Each subject was confronted with three sets of 50 word-pairs. Order of presentation of the three sets was counterbalanced within each group of subjects.

Each set of 50 word-pairs consisted of 25, randomly selected, in which the two words meant the same thing and 25 in which they meant different things. Incorrect pairing was done by random assignment of the remaining words after the matched pairs had been selected. Separate randomizations were used for each language combination, although the same randomizations were used across methods of presentation.

Subjects were 269 undergraduates in an elementary psychology course. Eighteen were eliminated on the basis of knowledge of one or more of the foreign experimental languages, and eleven more were eliminated at random in order to equalize the size of all subject groups ($N = 40$ per group). All subjects were aware of the general purpose of the study and were informed that in only 25 word-pairs in each set of 50 did the two words mean the same thing. To increase involvement in the task, subjects were told that above-chance accuracy in guessing had already been demonstrated and that on completion of the task they could correct and score each other's papers. (Scoring was checked independently by the experimenters.)[1]

RESULTS

For each subject three scores were available, one for each of the three sets of 50 word-pairs. Scores were the number of correct guesses out of fifty. Two pseudo three-way analyses of variance were made of the data, one for the English-Foreign condition and one for the Foreign-Foreign condition. In each, individuals formed the rows, language combinations the columns, and methods of presentation (A, AV, and V) the blocks.

The results of the analysis for the English-Foreign language combinations are given in Table 1. Differences among languages were quite significant ($p < .001$) while differences among methods of presentation bordered on significance ($p < .05$). The differences among languages were determined by the relative ease with which the subjects guessed the equivalence of the English and Hebrew words, and among methods of presentation, by the superiority of the A-V and V methods to auditory presentation.

Table 2 presents the results of the analysis for the Foreign-Foreign language combinations. Again differences among languages were significant ($p < .001$) but those among methods of presentation were not. All differences among languages were significant. In this analysis there was a highly significant interaction effect with C-J declining in accu-

[1] Two additional tables (*a*) a seven-page table giving the six sets of 50 word-pairs and the number of 40 subjects marking each word-pair as meaning the same thing under the three methods of presentation, and (*b*) a four-page table containing the list of artificial words and the frequency of selection of alternative response choices by 49 subjects have been deposited with the American Documentation Institute. Order Document No. 5093 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

racy with the addition of visual cues and C-H increasing in accuracy.

Tables 3 and 4 give the percentage of correct judgments for the several language combinations under the different methods of presentation. In the English-Foreign comparison, only E-H among the language pairs was guessed more accurately than chance, over all; among methods of presentation, both AV and V methods produced over-all above-chance guesses. In the Foreign-Foreign comparison, when methods of presentation are combined, the percentage of correct guesses for all languages is significantly different from chance, but in the C-H word-pairs the subjects would have done better if they had merely tossed a coin for their decisions. For methods of presentation, only the Audio form produced over-all above-chance results.

As a final analysis, the grand totals of correct judgments under the two conditions (E-F and F-F) were compared. The number of correct judgments made in the Foreign-Foreign comparisons was not significantly greater than that made in the English-Foreign.

These results leave the answer to the first question raised—i.e. the effect of method of presentation on accuracy of guessing—somewhat uncertain. In the E-F comparison visual cues tend to raise the number of correct guesses; in the F-F comparison the effect of adding visual cues is variable depending upon the language pairs.

The second question, i.e. can subjects pair words from two different languages with above-chance accuracy when the sample of words is a random one from a specified population, can be answered affirmatively—for some language pairs. There was a wide range in guessing accuracy for the six language combinations with Chinese-Japanese word-pairs being matched most accurately and Chinese-Hebrew word-pairs the least accurately. In general, the significance levels for all comparisons are much lower than previously reported.

The third question, i.e., what are the cues that lead to correct guessing, required an additional analysis of the data. It is apparent that above-chance accuracy is determined by the fact that, on the average, the agreement among the subjects on the sameness or differentness of the word-pairs tended to correspond to objective sameness or differentness in word meanings. Consequently, it was possible to approach the problem of the cues that lead to accurate guessing by studying the factors that led to agreement among the subjects, whether or not such agreement was in fact correct on any particular word-pair.

There are undoubtedly many factors that lead to intersubject agreement. However, the most promising set of determinants for analysis lies in the objective stimulus characteristics of the word-pairs. To determine these cues, the individual responses of each group of 40 subjects for each language combination were tabulated and those word-pairs isolated on

TABLE 2

VARIANCE TABLE FOR THE MATCHING OF WORDS IN TWO FOREIGN LANGUAGES

Source	*df*	Sum of Squares	*F*	*p*
Individuals	117	1271.3		
Languages	2	691.3	31.83	$<.001$
Method of presentation	2	6.7	.31	$>.10$
Languages × methods	4	145.8	4.02	$<.01$
Remainder	234	2122.3		
Total	359	4237.4		

TABLE 3

PERCENTAGE OF CORRECT JUDGMENTS OF THE EQUIVALENCE OF MEANINGS OF ENGLISH AND FOREIGN WORDS

Languages	Method of Presentation			Total
	Auditory	Audio-visual	Visual	
English-Japanese	47.2*	52.2	51.4	50.3
English-Hebrew	52.6	53.3*	53.0*	53.0*
English-Chinese	49.0	49.9	50.8	49.9
Total	49.6	51.8*	51.7*	51.0*

* $p < .01$.

TABLE 4

PERCENTAGE OF CORRECT JUDGMENTS OF THE EQUIVALENCE OF MEANING OF WORDS FROM TWO FOREIGN LANGUAGES

Languages	Method of Presentation			Total
	Auditory	Audio-visual	Visual	
Chinese-Japanese	57.0*	53.8*	53.7*	54.8*
Chinese-Hebrew	46.8*	47.7	49.9	48.1*
Hebrew-Japanese	52.6	53.3*	50.9	52.3*
Total	52.1*	51.6	51.5	51.7*

* $p < .01$.

which more than 70 per cent of the subjects agreed in marking them as "same" or "different." Such high percentages of agreement were quite common, averaging more than 20 out of each set of 50 word-pairs. The list of word-pairs thus isolated was inspected for cues that might have led to the high degree of agreement observed. A number of possible factors was apparent; the five that seemed to predominate were:

1. Length—words containing the same number of letters and/or syllables tended to be marked "same" and those of differing lengths "different."

2. Vowels—words containing exclusively or predominantly vowels of the same group (*a-o-u* or *i-e*) tended to be marked "same" whereas when the vowels in the two words were from different groups, they were judged "different."

3. Consonants—words containing exclusively the same consonant types (sibilants, explosives, etc.) were marked "same." Differing consonant types in the two words elicited the "different" response.

4. Hyphenation or spacing—where perceptible spacing occurred in both words they tended to be marked "same."

5. Connotation—if the two words or parts thereof suggested a single English concept, the two words tended to be marked "same."

These factors were not all equally frequent nor present to equal extent in the differing language pairs. Connotation, for example, appeared to be much stronger in the English-Foreign comparisons than in the Foreign-Foreign. Similarly, certain of the factors appeared to have considerable strength in eliciting a "same" response but did not affect the "different" judgment as markedly. The converse was also true.

In order to determine the relative potency of the postulated factors, the original six lists of 50 word-pairs were checked for the occurrence or nonoccurrence of the factors in each word-pair. Where a factor was present in a given word-pair and the largest frequency of response was in the direction expected, a plus tally was made for that factor. If the largest frequency of response was in the direction opposite that expected, a minus tally was made. A comparison of the proportion of plus tallies to total tallies for each factor gave the following potency ranking: connotation, vowels, length, spacing, and consonants.

Since the analysis described above was highly subjective, a second experiment was conducted to test the postulated effects of three of the factors described above (connotation, length, and vowels.) To control the occurrence of the factors, artificial "languages" were used, the words of which were constructed from nonsense syllables. Presentation of the words was visual only with each item consisting of one stimulus "word" paired with two response "words." The subject's task was to select that one of the alternatives which meant the same thing as the stimulus "word."[4]

The alternative choices in each item were constructed so as to match or to differ from the stimulus word in one or more of the three factors under consideration. Table 5 indicates the factors in which the two response words matched the initial word, the number of items of each type, an example of each type of item, and the percentage of 49 subjects who selected the predicted alternatives in each type of item. The last column indicates the rank order of potency predicted on the basis of the original derivation of the factors. Two rank orders of potency were predicted, one where one factor at a time was investigated while the others were held constant, and one for the factors pitted against each other.

In the construction of the words, nonsense syllables of low association value were used where connotation was to be minimal. For the words with high connotation value, nonsense syllables of high association value were used in part as well as some nonsense words constructed from the Kent-Rosanoff list of high associative stimulus words. In words in which vowels were matched, each vowel within one group (*a*, *o*, *u*, or *i*, *e*) was matched with itself and with its fellow members an equal number of times. In the words in which vowels did not match, *i* and *e* were paired equally frequently with *a*, *o*, and *u*. In the response words where length was to be different from the initial word, the response word was equally frequently longer and shorter. Differences in length were varied from two to five letters.

The data in Table 5 show one reversal from the predicted rank order of potency. The factor

[4] See Footnote 3.

TABLE 5

FACTORS ON WHICH RESPONSE "WORDS" WERE MATCHED WITH STIMULUS "WORDS" AND PERCENTAGE OF 49 *Ss* SELECTING THE FIRST ALTERNATIVE

(Artificial "words" formed from nonsense syllables)

Response Words		No. of Items	Example		Percentage of Subjects Selecting A	Predicted Rank Order of Potency
A	B					
Vowels Length Connotation	Vowels Length	9	famm	hoam noax	80	1
Vowels Length	Length	30	jiq	yil yom	66	2
Vowels Length	Vowels	9	wog	zos zotyom	71	3
Vowels Length Connotation	*None*	9	hur	wom wibpid	85	1
Vowels Length	*None*	9	yod	zoj xezhief	82	2
Vowels Connotation	Length	8	wat	lakcull jid	65	3
Length Connotation	Vowels	9	hevv	wayt meijzieb	66	4
Length	Vowels	8	wez	gebeviff xag	57	5
Vowels Length	Connotation	9	darq	neitz naox	56	6

of length turned out to be more effective in determining the subjects' decisions than was the factor of vowel similarity. All the percentages of agreement obtained were significantly greater than chance with the exception of that obtained in pitting C against L plus V. These results support the conclusions drawn from the empirical item analysis of the natural languages. Similarity in length of words, vowel groups, and connotation significantly influence the decisions of subjects as to similarity of meaning.

DISCUSSION

Certain discrepant results appear among the two studies cited and that reported here. Brown, Black, and Horowitz found that their subjects were able to guess the English meanings of Chinese words significantly above chance, whereas we did not. Maltzman, Morrisett, and Brooks report that their subjects were able to guess the English equivalents of Japanese words in a similarly successful fashion; we found that this was not the case for our subjects. There appear to be three possible sources for this disagreement: (*a*) method of presentation (audio, audiovisual, or visual), (*b*) design of stimulus presentation (simple paired comparison, single stimulus—double response choice and double stimulus—double response choice), and (*c*) the word sample used. The first source appears to be ruled out, for where we used the method of presentation employed by the other investigators, negative results were still obtained (Table 3). On the basis of the pilot study briefly mentioned, the difference between designs of stimulus presentation likewise did

not appear to be crucial. The third alternative thus remains as the most likely one although an interaction effect between (*b*) and (*c*) cannot be ruled out. According to this interpretation, it is possible to match words for meaning between English and certain foreign languages with better than chance success, without knowledge of the foreign equivalents. The process is facilitated by marked contrasts of meaning such as occur in antonym pairs or contradictory word pairs. Where a random sample of words from a general concept population is used and where meaning contrasts are not as great, correct guessing becomes much more difficult although it may still occur with better than chance success in some languages, e.g., English-Hebrew.

Maltzman, Morrisett, and Brooks also found that their subjects were unable to match Japanese-Croatian words of the same meaning whereas in the present study the subjects were able to pair Chinese-Japanese words and Hebrew-Japanese words significantly better than chance. However, our subjects were significantly poorer than chance on Chinese-Hebrew words. The conclusion of Maltzman *et al.*, that their results tend to disprove the general hypothesis of universal phonetic symbolism, would seem to be correct, but the positive findings observed raise certain interesting alternatives. The observation that noncognate languages may employ similar patterns of sounds (and length) to designate similar meanings—at least among frequently used words—offers an intriguing way of grouping languages. However, it should be emphasized that the restricted nature of the population of concepts used in this study does not permit conclusions about the degree of similarity between *total* word populations of any two of the languages used.

SUMMARY

Six groups of 40 English-speaking subjects were presented with the task of judging the equivalence of meaning of three sets of 50 pairs of words presented auditorily, audiovisually, or visually. The words were a random sample of concepts of high frequency of usage and the three sets were composed of pairs of words in English and Chinese, English and Japanese, and English and Hebrew, or Chinese and Japanese, Japanese and Hebrew, and Chinese and Hebrew.

The subjects were able to guess the English equivalents of Hebrew words and to pair Japanese and Chinese, and Japanese and Hebrew words at better than chance levels. In pairing Chinese and Hebrew words the number of correct guesses was significantly below chance. In the English-Foreign comparisons, visual presentation of the romanized versions of the foreign words tended to increase the percentage of correct guesses; but not in the Foreign-Foreign comparison. A significant interaction effect between languages and method of presentation was noted in the latter situation.

An item analysis of the subjects' responses indicated that similarity of word lengths, vowels, consonant types, spacing of compound words, and connotation markedly influenced the agreement among the subjects as to the sameness of meaning of word pairs. A second experiment indicated that agreement among subjects as to the meaning of synthetic words could be systematically varied by manipulation of the cues of length, vowels, and connotation.

The results of the study are interpreted as contrary to the hypothesis of a universal phonetic symbolism.

REFERENCES

1. BROWN, R. W., BLACK, A. H., & HOROWITZ, A. E. Phonetic symbolism in natural languages. *J. abnorm. soc. Psychol.*, 1955, **50**, 388–393.
2. EATON, H. S. *Semantic frequency list for English, French, German, and Spanish.* Chicago: Univer. of Chicago Press, 1940.
3. MALTZMAN, I., MORRISETT, L., & BROOKS, L. O. An investigation of phonetic symbolism. *J. abnorm. soc. Psychol.*, 1956, **53**, 249–251.
4. OSGOOD, C. E. & SEBEOK, T. H. (Eds.) Psycholinguistics. A survey of theory and research problems. *J. abnorm. soc. Psychol.*, 1954, **49** (4, Pt. 2—Suppl.).
5. TSURU, S. & FRIES, H. S. A problem in meaning. *J. gen. Psychol.*, 1933, **8**, 281–284.

Received May 31, 1956.

PERFORMANCE ON THE BENDER UNDER HYPNOTIC AGE REGRESSION

HAROLD B. CRASILNECK AND CARMEN MILLER MICHAEL

Southwestern Medical School, University of Texas

CONSIDERABLE debate may be found in psychological literature concerning the phenomenon called "hypnotic age regression." Some reports (1, 4, 5, 6, 7, 8, 11, 12, 14) have indicated that a subject can be taken back hypnotically to a given age and that his behavior then is exactly as it was at that age—that there is a genuine regression involving the total organism. Other studies (2, 3, 9, 10, 13, 15, 16, 18) suggest that hypnotically regressed subjects attempt to play an assigned role and to cooperate with the hypnotist's suggestions, but do not actually, organismically reproduce the earlier behavior. The present study addresses itself to this debate.

In the various experiments that have been designed to test the "role-playing" versus the "true regression" hypothesis, subjects have been given a variety of physiological and psychological tests. In such studies, particularly when psychological tests are used as a criterion of regression, the subject should not be aware of how an individual at the assigned age should behave. The writers believe that the Bender Motor Gestalt Test is one which allows for less of such awareness than has usually been the case in tests used in previous studies.

In the present experiment, the problem was to determine the absolute and relative levels of maturity of behavior of subjects under four kinds of conditions, and to test especially the hypothesis that under hypnotic age regression, subjects behave exactly as they did at a given, suggested age. For the reason previously mentioned, the Bender Motor Gestalt Test was selected as the instrument to be used in judging maturational level.

METHOD

Subjects

The *S*s were ten white female student nurses who had volunteered their services for this experiment. They ranged in age from 19 to 22 years with a mean of 20. Although the *S*s knew that the experiment "had something to do with hypnosis," they were naive concerning the specific problem involved. In each case *S* was repeatedly hypnotized over a period of time until it was definitely ascertained that she was a somnambule, capable of manifesting those signs described by Wolberg (17). Prior to the experimental procedure either the Wechsler-Bellevue Intelligence Scale (Form I) or the Wechsler Adult Intelligence Scale was administered by one of the staff psychologists not associated with the hypnotic sessions.[1]

Procedure

The Bender was given to each *S* individually under four conditions in the following sequence: (*a*) awake, (*b*) awake with instructions to pretend to be four years old, (*c*) hypnotized and instructed to pretend to be four, and (*d*) hypnotically regressed to four.

Condition 1. After entering the experimental room, *S* was seated at a desk, handed a pencil and a sheet of paper, and given the following instructions: "I have some cards like this and I want you to draw some things for me. Now make one like this and when you are through, I'll give you the next one." After completing the Bender, *S* was asked to put her name on the paper.

Condition 2. The *S* was then told, "Now I want you to pretend that you are four years old. I want you to do anything that is asked of you just like you would have done it when you were four years old." The *S* was again handed a pencil and paper, given the Bender, and asked to put her name on the paper.

Condition 3. Following this, *S* was hypnotized by the visual fixation technique into a somnambulistic state. She was then told that, "In a moment you will be able to open your eyes; however you will still remain deeply sound asleep. Upon opening your eyes, I want you to pretend that you are four years old. I want you to do anything that is asked of you just like you would have done it when you were four years old." Upon opening her eyes, *S* was again given the Bender and asked to put her name on the paper.

Condition 4. The *S* was then told to close her eyes and was hypnotically regressed to four years of age. During the regression, she was asked questions concerning her age, background, and professional knowledge. After it was reasonably established that *S* was manifesting the usual, expected signs of regression to four years of age, it was suggested that she would open her eyes, but remain sleeping, and that she would do what was asked of her. The Bender was then administered and *S* was told to put her name on the paper. The *S* was then hypnotically returned to her present age with amnesia for the entire episode.

Judgments of Maturational Level

Three clinical psychologists who were completely naive as to the nature of the experiment served as

[1] The authors wish to thank Dr. Arthur Lee Foster and John Gladfelter.

judges.[2] The 40 Bender protocols were disguised with regard to identity and arranged in a random order. They were given to the judges individually with the following instructions:

> Here are 40 Bender Gestalts. They are numbered from 1–40. You will notice that each is on a separate sheet of paper, except for #28 and #34, each of which is on two sheets.
>
> Each Bender has a name and a date and some other information on it. All of this has been covered over by some heavy blue paper. It is essential that you leave the blue paper in place and not look at the information underneath.
>
> You are requested to make a judgment concerning each of these Benders. Using your past experience and the help of Plate 62 in Bender's Monograph, estimate the maturation level in the visual motor gestalt function of each Bender. Each judgment should be one of the following: 3½ (under 4 years), 4, 4½, 5, 5½, 6, 6½, 7, 7½, 8, 8½, 9, 9½, 10, 10½, 11, or 11½ (over 11 years).
>
> Please use the attached Judgment Sheet to record your judgments. Your comments are not being solicited, but if you have any, you may make them in the space provided.

The protocols were then rearranged into ten sets. Each set consisted of the four records from each *S* in a random order. These materials were again given to the judges individually with the following instructions:

> Here are 10 sets of Bender Gestalts. There are four Benders in each set. Each set is from a different subject, but all four Benders in each set are from the same subject. Each Bender is on a separate sheet of paper, except for #28 and #34, each of which is on two sheets.
>
> Each Bender has a name and a date and some other information on it. All of this has been covered over by some heavy blue paper. It is essential that you leave the blue paper in place and not look at the information underneath.
>
> The Benders are numbered randomly from 1–40. Each Bender also has a name (initials). Of course, in each set of four all the Benders will have the same name.
>
> Your are requested to rank the four Benders in each set according to your estimate of the relative maturational level in the visual motor gestalt function. Please rank as "a" the one which you think is most mature, "b" the next most mature, "c" the next, and "d" the one which you think is most immature. Use your past experience and the help of Plate 62 in Bender's Monograph.
>
> Please use the attached Judgment Sheet to record your judgments. Your comments are not being solicited, but if you have any, you may make them in the space provided.

The reliability of the judges' absolute ratings of the maturational level of the 40 Benders was determined by calculating the intercorrelations of judgments by the method of rank-difference. The correlation coefficients were .90, .84, and .89. Since the reliability of the judges could be regarded as high, their absolute judgments were averaged for each case and their composite ratings used for the analysis of the data on maturational level.

[2] The authors wish to express gratitude to Mrs. Genette Burruss, Dr. Rosaline Goldman, and Dr. Harold Winer.

RESULTS

The mean maturational level of each of the four conditions under which each *S* took the Bender was determined. The means for Conditions 1 through 4 were, respectively, as follows: 11.2 years, 9.9 years, 7.8 years, and 7.3 years.

The significance of the difference between any two of the above means was determined by computing the significance of the difference between correlated means. All except two of the differences were significant at the .01 level of confidence. The difference between Mean 1 and Mean 2 was significant at the .02 level. The difference between Mean 3 and Mean 4 was not significant. When the significance of the difference between each of the four means and the hypothetical mean of four years was assessed, all differences were significant at the .01 level.

On the basis of the *S*s' Wechsler-Bellevue and WAIS IQ scores, estimates were made of their probable mental ages when they were chronologically four years old. The Wechsler scores were translated into Standard scores and then converted into comparable Stanford-Binet IQs for four-year olds. The mean estimated mental age thus obtained was 4.6 years. The difference between each of the means for Conditions 1 through 4 and this index of mental age (4.6 years) was determined by computing the significance of the difference between correlated means. Each of the four differences was significant at the .01 level of confidence.

These results with absolute ratings of maturational level led to the following question: If the judges make *relative* ratings, i.e., if they rank each of the four Benders from the same *S* in order of maturity, will the difference between the rankings under the different conditions, especially between Conditions 3 and 4, be significant? The judges' rankings were averaged for each case, and the significance of differences was determined by the method of paired replicates. The findings were that the difference between rankings under any two conditions was significant at the .01 level *except* the difference between the third and fourth conditions, which was not significant. The means of the judges' rankings from most to

least mature under the four conditions were 1.1, 2.0, 3.3, and 3.6, respectively.[2]

DISCUSSION

The foregoing results indicate that when *S*s are asked to perform at a given lower age level, they comply to a degree but do not seem to reach the actual age level suggested. They behave more nearly like the given age when the suggestion is made when *S* is under hypnosis than when in a waking state. When in a state of hypnotic age regression, some of the *S*s tended to behave more nearly like the suggested age than they did while hypnotized but merely instructed to pretend they were the suggested age. However, the difference in performance under these two conditions was not found to be consistent or statistically significant. It appears, then, that the results do not support the idea that hypnotically regressed *S*s behave exactly as they did at the suggested age.

Since all *S*s were given the Bender under four conditions in the same sequence, it is possible that they may have learned by practice and that therefore their performance level was not as low as it would have been had they not had the practice. Even if this is true, however, it would still support the inference that a total regression does not take place.

One may, of course, question the authenticity of the hypnotic technique and the criteria which were used as evidence that the *S*s were actually in a somnambulistic state and capable of demonstrating regression as well as could be expected. Each of the *S*s manifested those signs of somnambulism and regression listed by Wolberg (17), and during the process of age regression, all *S*s manifested personality changes of a regressive nature. For example, handwriting, pencil grip, vocabulary, and pronunciation of words became more and more childish. When hypnotically regressed to the age of four, one *S* began to cry when she was urged to put her name on the paper. When questioned she said she was crying because she did not know how to write her name. Four *S*s, who did not ordinarily bite their nails, began to do so when hypnotically regressed to an early age. Still other signs, usually accepted as evidence of age regression, were obtained after the actual experimental data were collected. For example, while still in a state of hypnotic age regression, the *S*s were given the suggestion that they would return to the age of one year. They then no longer answered when asked questions. Some *S*s were further regressed to four months of age at which time the sucking reflex was elicited.

Although the Benders obtained while the *S*s were hypnotically regressed (to the age of four years) were reliably rated by the judges to reflect an average maturational level of 7.3 years, it should not be concluded that their performance was typical for the age of seven or that an organismic regression down to seven years actually occurred. All of the judges spontaneously commented that it was difficult for them to assign an absolute maturational level to the Benders because there was such marked intratest variability on most of them. One of the judges rated each gestalt separately and then took the median rating as the best approximation of over-all functioning level. Another judge reported that, "Some of the subjects were children; some were adults; some were psychotic; and some were organic." The judges also indicated that it was much easier to make relative judgments than absolute judgments.

Although the results indicate that hypnotically regressed *S*s do not actually behave totally and exactly as they did at the suggested age, one may legitimately speculate about what might account for the behavioral changes which do occur. The *S*s appear to be attempting to follow the hypnotist's suggestions as exactly as they can, and hypnotic techniques seem to increase their desire or to facilitate their attempts to follow suggestions, or, as some investigators have called it, to play an assigned role. But in addition to the probable role-playing aspects of behavior under hypnosis, the possibility still exists that fragments of earlier behavior may be reproduced during hypnotic regression. There may occur, for example, a reinstatement of earlier reflex or conditioned reflex behavior even though a total regression does not occur.

There are two methodological issues, however, which limit the interpretation to be placed on the present data. First, maturational level was defined in terms of judges' ratings of

[2] The authors are indebted to Dr. Alvin J. North for his help in setting up the statistical design of this research.

Bender protocols. Like all ratings, these estimates may have been subject to the error of central tendency, the probability that judges will use the middle of the range rather more than the extremes. Since the range of possible mental age ratings ran only from four to eleven, this error of central tendency, if it actually occurred, could conceivably account for the significant differences between the obtained values and the hypothetical base of 4.6 years. Only the absolute judgments, of course, would be affected, the relative comparisons being free from any influence from this source.

Second, what has been said here assumes that ratings by qualified judges of Bender protocols constitute a legitimate and useful operational definition of maturational levels. Other defining operations might well lead to rather different findings. Nevertheless, the present procedure seems to be strongly defensible on at least two counts: (*a*) the high reliability of the judges indicates at least a great deal of commonality in clinical opinion about maturational level as revealed by the Bender, and (*b*) *Ss* are probably less able to imitate, either consciously or unconsciously, their performance at presumably lower age levels on the Bender than is the case with other indices of maturational development. Consequently, it would appear that the method used was a sound one in the light of the hypothesis tested.

SUMMARY

In this study, the problem was to determine the absolute and relative levels of maturity of behavior of *Ss* under four conditions, and especially to find whether hypnotically regressed *Ss* behave exactly as they did at a suggested age. The Bender Motor Gestalt Test was selected as the instrument to be used in judging maturational level. Ten *Ss* were administered an intelligence scale. These *Ss* then took the Bender under the following conditions: awake, awake and pretending to be four years old, hypnotized and pretending to be four, and hypnotically regressed to four. Independent ratings by three judges revealed, among other things, that when *Ss* are asked to perform at a given lower age level, they comply to a degree, but do not seem to reach the actual age level suggested.

REFERENCES

1. BERGMAN, M. S., GRAHAM, H., & LEAVITT, H. C. Rorschach exploration of consecutive hypnotic chronological age level regressions. *Psychosom. Med.*, 1947, **9**, 20–28.
2. FARBEROW, N. L., & SARBIN, T. R. A clinical study of role and self in hypnotic age regression. *Amer. Psychologist*, 1950, **5**, 305. (Abstract)
3. KEIR, G. An experiment in mental testing under hypnosis. *J. ment. Sci.*, 1945, **91**, 346–352.
4. KLINE, M. V. Hypnotic age regression and intelligence. *J. genet. Psychol.*, 1950, **77**, 129–132.
5. KLINE, M. V. Hypnosis and age progression: a case report. *J. genet. Psychol.*, 1951, **78**, 195–206.
6. KLINE, M. V. Childhood fears in relation to hypnotic age regression: a case report. *J. genet. Psychol.*, 1953, **82**, 137–142.
7. KLINE, M. V. Hypnotic retrogression: a neuropsychological theory of age regression and progression. *J. clin. exp. Hypnosis*, 1953, **1**, 21–28.
8. LEEDS, M. An hypnotic regression series. *Persona*, 1949, **1**, (1), 13–16.
9. MCCRANIE, E. J., CRASILNECK, H. B., & TETER, H. P. The electro-encephalogram in hypnotic age regression. *Psychiat. Quart.*, 1955, **29**, 85–88.
10. ORNE, M. T. The mechanisms of hypnotic age regression. *J. abnorm. soc. Psychol.*, 1951, **46**, 213–225.
11. PLATONOW, K. I. On the objective proof of the experimental personality age regression. *J. gen. Psychol.*, 1933, **9**, 190–209.
12. SARBIN, T. R. Mental age changes in hypnotic age regression. *Amer. Psychologist*, 1946, **1**, 456–457. (Abstract)
13. SARBIN, T. R. Mental age changes in experimental regression. *J. Pers.*, 1950, **19**, 221–228.
14. SPIEGEL, H., SHOR, J., & FISHMAN, S. An hypnotic ablation technique for the study of personality development. *Psychosom. Med.*, 1945, **7**, 273–278.
15. TAYLOR, A. The differentiation between simulated and true hypnotic regression by figure drawings. Unpublished master's thesis, City Coll. of New York, 1950.
16. TRUE, R. M. Experimental control in hypnotic age regression states. *Science*, 1949, **110**, 583–584.
17. WOLBERG, L. R. *Medical hypnosis*. Vol. 1. New York: Grune & Stratton, 1948.
18. YOUNG, P. C. Hypnotic regression—fact or artifact? *J. abnorm. soc. Psychol.*, 1940, **35**, 273–278.

Received May 28, 1956.

DECISION PROCESSES IN COMMUNICATION NETS[1]

MARVIN E. SHAW, GERARD H. ROTHSCHILD, AND JOHN F. STRICKLAND

The Johns Hopkins University

THERE is ample experimental evidence that the communication structure of a group influences the performance and satisfaction of the group when the task is to solve problems having logically imperative solutions (2, 8, 10, 12, 14, 15). However, there seems to be no evidence that the same kinds of effects obtain when the problem to be solved does not have a generally agreed upon solution, e.g., a problem requiring agreement among the group members about the proper way to deal with an issue involving interpersonal relations. The first experiment reported in this article attempted to determine the effects of certain communication nets upon group performance and satisfaction when the task is to solve such a "human relations" problem. Experiments II and III were attempts to answer certain questions raised by the results of Experiment I.

EXPERIMENT I

Method. The apparatus used in this experiment is the same as that described in earlier reports (14, 15). It consists of four cubicles, each of which has a worktable, work materials, and a switch which controls a light and timer located at *E*'s position. The cubicles are interconnected by means of slots through the walls separating them. Group members can communicate with each other by writing messages on cards and passing these cards through the slots. Various communication nets can be arranged by closing the necessary slots. The communication nets investigated in this study are shown in Fig. 1.

The *S*s for this experiment were students (all male) at The Johns Hopkins University. All were highly experienced, having served previously as *S*s in this kind of experimental situation for a total of 12 sessions during which they solved a total of 24 problems involving simple arithmetical manipulations. For purposes of counterbalancing, the experimental design required that eight groups of four *S*s be assigned to each of the three communication nets. One group which had been assigned to the "slash" failed to report for the experimental session, and since no other experienced *S*s were available the data reported here are for eight groups in the "star," eight in the "comcon," and seven in the "slash." It appears highly unlikely that failure to fill this one cell in the design resulted in any serious bias in the final results.

Two human relations problems were used. One was a modified version of Deutsch's "The World War II Veteran" problem (3). As presented to the *S*s in this experiment, the central figure in the story was an executive who had had an affair with his secretary. The secretary had subsequently resigned to get married. The executive's wife is intelligent but suffers from a traumatic childhood experience which made her especially sensitive to this kind of unfaithfulness. The question: What should the executive do, assuming that for his own peace of mind he cannot just forget the matter?

The other problem was essentially the same as that used by Festinger and Hutte (6). The central figure in this story is Henry, the son of a physician, who has a friend, Jim, who is incurably ill. Both are in love with the same girl. Just as Henry has decided to ask the girl to marry him, she announces her engagement to Jim. The girl does not know that Jim is incurably ill, although Henry does. The question: What should Henry do and say?[2]

Except for the communication net and the order in which the problems were presented, the procedure was the same for all groups. The *S*s were seated in their cubicles and handed a mimeographed set of instructions. The *E* then read the instructions while the *S*s followed along on the printed page. The instructions informed *S*s: (*a*) of the nature of the communication net (which was, with minor exceptions, the same as during previous sessions), (*b*) that they were to solve problems having no correct or incorrect answer, but which merely called for a group decision, (*c*) that a unanimous decision was desirable but that they could give a minority report if they were convinced that unanimity could not be attained, and (*d*) that they would be given 20 minutes to discuss the problem, but could signal *E* by throwing their switches if agreement was reached before the end of this period. The lack of emphasis upon time is in contrast to instructions given in previous investigations. Problems were presented in counterbalanced order.

After both problems had been considered and decisions rendered, *S*s were required to indicate on an 11-point scale how well they enjoyed their task in the group and how they would rate the performance of their group.

Results. Of chief interest were the differences due to the communication net and to positions within nets. In addition, data were evaluated

[1] This experiment was done under Contract N5-ori-166, Task Order 1, between the Office of Naval Research and The Johns Hopkins University. This is Report No. 166-I-198, Project Designation No. NR 145-089, under that Contract.

[2] The second problem appeared to be much more interesting to the *S*s in this experiment, perhaps because most were unmarried and hence the situation tapped a more relevant frame of reference.

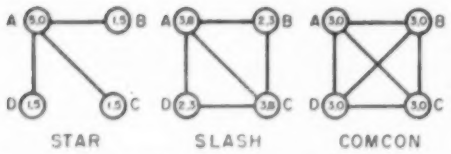

Fig. 1. The Experimental Nets

Circles represent persons and the lines represent two-way communication channels between persons. Numbers within circles are the Independence scores (13) for the positions within the net.

TABLE 1

Mean Number of Communication Units Transmitted by *Ss* in Different Positions

Nets	Positions				p
	A	B	C	D	
Star	50.8	22.8	18.1	21.8	$< .01$
Slash	41.6	29.6	38.3	29.0	ns
Comcon	31.5	33.3	32.8	32.6	ns

for differences due to trials (first vs. second problem) and to problems. In agreement with the results of previous studies, trials differed significantly with respect to both time and message units required to reach a decision ($p < .01$). This finding is a general one and will not be referred to again. Differences due to problems were not significant in any of the analyses performed.

There were marked differences among the groups in different communication nets in mean times required to reach a decision (16.62 min. in the star, 14.00 min. in the slash, and 9.71 min. in the comcon). The F test revealed that these differences are statistically reliable ($p < .01$). This finding is in agreement with the results obtained when the problems to be solved were of the arithmetical type, but it might be noted that the magnitudes of the differences are somewhat greater than those typically reported in earlier investigations.

There were no significant differences among positions within nets with respect to time scores. In fact, under the conditions of this experiment it would have been very surprising if such differences had been found.

Content of messages was analyzed into units by the method developed in earlier studies (12, 14, 15) according to which a message unit is defined as any simple sentence or any meaningful part of a complex or compound sentence. There were no significant differences among nets in number of message units required to reach a decision (means were 28.4, 34.6, and 32.6 for the star, the slash, and the comcon, respectively). This does not agree with the findings of earlier studies, but this failure seems to be due almost entirely to the great differences in time required to reach a decision. If message units per unit time are considered, nets are ordered in the same way as in other experiments. (Mean message units per min. were 3.60, 4.95, and 7.19 for the star, the slash, and the comcon, respectively. Differences are statistically reliable [$p < .001$].)

Mean number of message units transmitted by *Ss* in the various positions within nets are shown in Table 1. Only the differences in the star are significant, but the differences in the slash are in the same direction as that found in earlier studies. That is, *Ss* in the more centralized positions sent more messages than did *Ss* in the more peripheral positions. It is also interesting that the number of messages directed toward deviators (i.e., those who disagreed with the majority opinion) was considerably greater than the number directed toward nondeviators in equivalent positions (mean number message units received by deviators: 11.9, 19.3, and 22.7 in the star, the slash, and the comcon, respectively; by nondeviators: 9.2, 14.3, and 16.9 in the star, the slash, and the comcon, respectively). Differences between deviators and nondeviators were significant at well beyond the .01 level of confidence in each of the three nets. This finding agrees with results obtained under very different conditions by Festinger and Thibaut (5) and by Schacter (11).

There were no significant differences among nets or among positions within nets either in ratings of satisfaction or in ratings of performance. However, it is interesting that the direction of the differences in ratings of satisfaction were opposite to that found in previous investigations (mean ratings of satisfaction: 9.09, 8.86, and 8.63 for the star, the slash, and the comcon, respectively; in the star, mean ratings of satisfaction were 8.3 for Position A and 9.4 for Positions B, C, and D, where higher numbers indicate greater satisfaction). A possible explanation for this result is given in a later section.

The frequency with which individuals disagreed with the majority opinion was very small in all three nets; hence, application of

statistical tests was not justified. However, it is interesting to compare the percentage of disagreements occurring in the three nets. The percentage of total opinions (decisions) which disagreed with the majority of the particular group was 18.75 in the star, 16.07 in the slash, and 9.38 in the comcon. Thus, despite the fact that *S*s in the star and the slash spent more time trying to reach an agreement, the degree of unanimity actually attained was less than in the comcon.

Differences in disagreements by *S*s in the various positions within nets are shown in Table 2. It can be seen that frequency of disagreement correlates rather well with the centrality of positions. In the star the most central position never disagreed with the majority opinion. Likewise the more central positions in the slash showed markedly fewer disagreements than did the more peripheral positions.[3]

Discussion. Taken together, the results given above present a reasonably consistent picture of the group process. The time required to reach a decision was longest in the star and shortest in the comcon, in agreement with earlier studies in which arithmetical problems were used. Number of message units required to reach a decision in the three nets did not differ significantly, but units transmitted per minute did. In other words, groups in the comcon reached a decision more quickly than did groups in the other two nets but did so by transmitting messages at a more rapid rate.

It has been suggested that the time differences are due entirely to the fact that in the star and in the slash the burden of writing falls mainly on one or two people; i.e., that differences are due entirely to the effects of the net upon ease of intragroup communication. Fortunately, this hypothesis can be checked rather easily. If it is true, then the rate of message transmission should be essentially the same for the *S* sending the greatest number of messages in each net. This is not the case; the star is significantly slower than either the slash or the comcon ($F = 6.85$, $p < .01$). Also, it would be expected that if we selected the one *S* in each group who sent the greatest number of messages for that group, and then compared nets with respect to number of messages transmitted by these *S*s alone, nets would differ significantly. This is not found ($F = 1.31$; $F_{.05} = 3.49$). Finally, if this hypothesis were true one would expect that time, measured from the "go" signal until the last *S* had thrown his switch, would correlate significantly with the greatest number of messages sent by any one *S* in that group. Again, the result is negative ($r = .097$; $r_{.05} = .396$). The results of these analyses make it clear that the interrelationships of nets, decision times, and messages are not simple ones.

TABLE 2
PERCENTAGE OF DECISIONS DISAGREED WITH BY *S*s IN DIFFERENT POSITIONS

Nets	Positions			
	A	B	C	D
Star	0	31.3	25.0	18.8
Slash	7.1	28.6	7.1	21.4
Comcon	6.3	12.5	6.3	12.5

Another finding that requires some discussion is the relationship between net and frequency of disagreement with the majority opinion. There are two possible explanations. One is that the *S*s in the more central positions were using their position to bring other members around to their own viewpoint. A second possibility is that *S*s in the more central positions who held deviant opinions were subject to more direct pressure than were those in peripheral positions; this added pressure caused them to change more quickly and more frequently than peripheral deviators. The latter interpretation also explains the fact that ratings of satisfaction by *S*s in different positions were not significantly different. The peripheral positions may have enjoyed being able to maintain a deviant opinion without being subjected to too much group pressure; on the other hand, *S*s in the more central positions may have had their satisfaction dampened by being vulnerable to group pressure.[4]

The results of Experiment I do not permit a choice between these two alternatives. Ex-

[3] The slight aberration in the comcon is due largely to the fact that one group split two-two; in this case, two positions were arbitrarily selected as being deviant.

[4] Another factor in the determination of ratings of satisfaction is saturation (8), i.e., the degree to which the channel capacity is reached or exceeded. When the communication demands upon persons in central positions become too great, their satisfaction with the position is reduced. The saturation point was reached in the star and to some extent in the slash.

periment II was designed to shed further light upon this question.

EXPERIMENT II

Purpose. This second experiment was designed to examine further the effects of the position within a communication net upon the ability of the *S* in the position to maintain an opinion in the face of opposition. Specifically, we wanted to know if an *S* in a central position would change his judgment of a relatively nonambiguous stimulus more readily than would an *S* in a peripheral position when faced with different judgments by fellow group members. If so, we would have reason to believe that proneness to change (tendency to conform) is related to degree of vulnerability to direct pressure. If the opposite should occur, i.e., if *S*s in central positions changed less readily than *S*s in peripheral positions, we would have reason to believe that central *S*s were using their favorable position to get others to agree with them. Also, analysis of message content would reveal this process if it should occur.

Method. Since we were interested primarily in comparing central and peripheral positions, we selected the net in which this differentiation is greatest, namely, the star shown in Fig. 1. The task was to judge the number of clicks which were presented to *S*s via tape recorders and earphones. Each cubicle was supplied with a headset, and a selector panel was set up so that each headset could be connected with either of two tape recorders. With this arrangement, all *S*s could be given the same number of clicks, or the number given could be different for different *S*s.

Three conditions were used. In Condition I the central person heard one number of clicks and the other three *S*s heard a different number of clicks; in Condition II the central *S* and one peripheral *S* heard one number of clicks and the other two *S*s heard a different number of clicks; and in Condition III one of the peripheral *S*s heard one number of clicks and all others heard a different number. .

The *S*s were 96 high school girls ranging in age from 17 to 19 years. *S*s were divided into 24 groups of four. All groups went through all conditions, but in counterbalanced order.

The clicks had the following characteristics: duration, 5 msec.; rate, 4 per sec.; and intensity approximately 55 db SPL. The number of clicks given on any one trial varied from four to nine. It has been shown (7) that under these conditions clicks can be counted with great accuracy (better than 95% correct). Discrepancies created by sending different numbers of clicks to different *S*s were either one or two. Following a familiarization trial in which all *S*s heard the same number of clicks, each group was given two trials in each condition, one with a discrepancy of one and the other with a discrepancy of two. No differences due to discrepancy were observed; therefore, the results from the two trials were combined for purposes of analysis.

Each *S* was required to make an independent estimate of the number of clicks heard on each trial after discussing it with others in the group via written messages. At the end of the experimental session each *S* was required to rate her satisfaction during each condition. On the rating scale, conditions were identified simply as the first one-third, the middle one-third, and the last one-third of the experimental session.

Results. The results of Experiment II are given in Table 3. These results were considered both in terms of the group as a whole, and in terms of different treatments of *S*s within groups. In the first kind of analysis it was possible to isolate the effects due to (*a*) conditions, (*b*) order in which conditions were presented to the group, and (*c*) the interaction of these two variables. In the second kind of analysis, scores for each condition were

TABLE 3
OUTCOME OF EXPERIMENT II

		Condition I		Condition II			Condition III		
		C*	P_{pp}	C_p	P_c	P_p	C_{pp}	P_{cp}	P
Change of estimate	Per cent	58.3	9.0	12.5	12.5	27.1	2.0	1.0	64.5
Time (min.) to decide	Mean	3.5	2.9	4.9	3.7	4.0	2.8	2.4	2.9
	SD	1.9	1.7	2.5	2.4	2.1	2.0	1.5	1.8
Message units per channel	Mean	9.1	10.0	12.9	15.7	15.1	8.5	8.6	9.9
	SD	6.6	6.6	5.4	7.9	7.0	4.5	4.2	5.0
Message units received	Mean	30.0	9.1	45.7	13.8	12.5	27.1	7.3	9.0
	SD	19.6	17.8	15.9	7.9	17.4	14.0	10.3	5.0
Ratings of satisfaction	Mean	6.0	6.0	6.1	6.1	5.9	6.4	5.9	5.7
	SD	0.9	0.8	0.8	0.7	0.8	0.6	0.9	1.1

* The capital letters "C" and "P" stand for central position and peripheral position, respectively; the subscripts "c" and "p" indicate position and number of supporters. For example, C indicates a central person with no support, P_{pp} indicates a peripheral person with the support of two other peripheral persons, etc.

analyzed separately since the different conditions determined the treatments of the *S*s.

The group analysis revealed significant differences due to conditions for time to reach a decision ($p < .001$) and for number of messages ($p < .001$); differences in frequency of change and ratings of satisfaction in the various conditions were not significant. Orders of condition presentation resulted in significant differences in frequency of change ($p < .01$), time ($p < .001$), and number of messages ($p < .01$), but differences in ratings of satisfaction were not significant. The condition × order interaction was significant for time scores ($p < .05$).

These results are apparently due to the fact that in Condition II (where the group was split two/two) changes of estimate were more likely to be in the nature of compromises than were the changes in the other two conditions. This procedure, however, required more time and messages. When Condition II came early in the experimental series, *S*s were "conditioned" to the compromising procedure, and also became less sure of their ability to count clicks accurately. Thus, more changes occurred when Condition II came in the first than in the second or third part of the experimental period. The significant interaction term with respect to time to reach a decision is due to the fact that this score decreased with practice.

The individual analysis revealed significant differences due to treatments for frequency of change in all three conditions ($p < .01$ to $p < .05$), for message units received in all three conditions ($p < .001$ in each case), for time to decision in Conditions I and II ($p < .001$, $p < .03$, respectively), and for ratings of satisfaction in Condition III ($p < .05$). No differences in message units per channel were found.

In both Condition I and Condition III the deviator changed her estimate much more frequently than did the nondeviators. The central deviator (C) did not differ from the peripheral deviator (P) in this respect. However, the central deviator apparently did spend more time trying to get the others to agree with her than did the peripheral deviator, and the fact that peripheral nondeviators in Condition I (P_{pp}) changed significantly more ($p < .05$) than did their counterpart in Condition III (P_{ap}) indicates that she may have had some measure of success. The effect of position within the communication net upon frequency of change of estimate is shown more clearly in Condition II in which both the central *S* and the peripheral *S* who agreed with her changed significantly less frequently ($p < .05$) than did the others in the group. It appears that the *S* in the more central position attempted to use her position to get others to agree with her estimate, but if the opposition became too great, she changed as frequently as did *S*s in the peripheral positions.[8]

Examination of the percentages given in Table 3 shows also that frequency of change of estimate is directly related to the number of persons who disagree with the estimate of a given *S* ($p < .001$). This result, of course, agrees with results reported by Asch (1).

The findings with respect to ratings of satisfaction are understandable if it be assumed that the increased satisfaction resulting from a favorable position in the net is counteracted by being in disagreement with others in the group. In Condition III these effects are both in the same direction and the expected significant differences among *S*s were found; in Condition I the two influences are in direct opposition and no differences were found; and in Condition II the issue is confused, but slight, insignificant differences in the expected direction were found.

EXPERIMENT III

Purpose. The results of Experiment II are reasonably straightforward, but there are some questions still unanswered. In the first place, perhaps the most interesting finding was that in Condition II both the central *S* and the peripheral *S* who agreed with her changed less frequently than did the other two *S*s in the group. But in this condition one supporter automatically meant two opposers, and vice versa. Thus, it is impossible to determine whether this finding is due to the fact that the central *S* had a supporter, or to the fact that she had only two opponents, or merely to the fact that the opposition was not unanimous.

In the second place, pressures exerted by other group members were permitted to vary

[8] One central *S* reported to the peripheral *S*s that others had agreed with her own estimate, when actually they had not. In this case, all the peripheral *S*s changed their estimate to agree with that of the central *S*.

with the personality of the *S*s; consequently, the amount and type of pressure actually exerted on the *S*s in different positions might have been affected by factors other than position within the net and/or the support-opposition ratio.

Finally, a trial varied in length so that the observed differences might have been due to differences in the amount of time spent trying to reach agreement. This factor might also have influenced the ratings of satisfaction. Experiment III was designed to alleviate these difficulties.

Method. The subjects for this experiment were male undergraduates at The Johns Hopkins University. Before the beginning of Experiment III, six groups drawn from this population were run under the conditions provided in Experiment II. Since their responses were indistinguishable from those of the high school girls, it was concluded that any differences between the results of Experiment II and Experiment III probably would not be due to population differences.

The communication net was again the star shown in Fig. 1, and the task was to estimate the number of clicks heard via earphones. In this experiment, however, *E* enlisted the aid of three confederates who ostensibly served as *S*s but made estimates according to *E*'s instructions. Thus, only one person (*S*) in each group was naive and it was his responses that were of interest. Each confederate was provided with a list of "agreeing" messages (e.g., "I also heard six clicks"), a list of "disagreeing" messages (e.g., "No, I heard five clicks"), and a list of "neutral" messages (e.g., "Sorry, I could not count them"). On critical trials, each confederate waited until an estimate had been received from *S*, and then he sent a message taken from the appropriate list. Further communications from *S* were answered by transmission of the next message on the list. This process continued until the end of the two-minute discussion period. Thus, the content of the messages sent by the confederates was controlled, but the number of messages sent was not. However, the variability within conditions was not great.

Six conditions were investigated involving three degrees of support-opposition in each of the two positions (central and peripheral) within the net. The three support-opposition conditions were designed to isolate the variables which were operating in Condition II of Experiment II. In one condition, one confederate agreed with *S*'s estimate and the other two were neutral. No change of estimate was expected in this condition, but it was included to obtain ratings of satisfaction and measures of message units transmitted. In the second condition, one confederate agreed and two disagreed with *S*, and in the third condition, two confederates disagreed with *S* and one was neutral.

Ten *S*s were run under each of the six conditions. Each group was given ten trials. On each trial *S*s were allowed two minutes for discussion after listening to the series of clicks and before making their estimates. On six trials, all confederates agreed with *S*, and on four, the critical trials, the confederates followed *E*'s instructions as described above. The discrepancy created by confederates disagreeing with *S* was always two. The maximum frequency of change for any one *S* was therefore four and the maximum amount of change (magnitude) was eight.

At the end of the experimental session each *S* was asked to rate his satisfaction with the group experience.

TABLE 4

OUTCOME OF EXPERIMENT III

		Central Position			Peripheral Position		
		ANN[a]	ADD	NDD	ANN	ADD	NDD
Frequency of change	Per cent	0	12.5	55.0	0	12.5	30.0
Magnitude of change	Mean	0	0.9	4.4	0	0.5	2.0
	SD	0	1.4	2.9	0	0.2	2.8
Message units per channel	Mean	8.2	13.2	11.9	7.3	8.2	8.0
	SD	3.6	3.6	3.4	3.3	2.2	3.9
Message units received	Mean	21.5	23.2	24.5	13.8	13.5	16.2
	SD	4.2	3.4	4.1	1.8	1.6	4.8
Ratings of satisfaction	Mean	8.8	8.7	8.6	7.8	7.3	5.8
	SD	0.4	0.7	1.0	1.5	2.2	2.1

[a] The letters "A," "D," and "N" stand for agreeing *S*, disagreeing *S*, and neutral *S*, respectively. Thus ANN means that one confederate agreed with the naive *S* and two gave no estimate, ADD means one confederate agreed and two disagreed, etc.

Results. The results of Experiment III are given in Table 4. Although the central *S* changed his estimate in the "two disagree-one neutral" condition more frequently than did the peripheral *S*, the over-all difference in frequency of change for *S*s in the two positions was not statistically reliable. The mean magnitude of change was greater for the central than for the peripheral *S* in both pressure conditions, and this difference was significant ($p < .05$). Likewise, the central *S* sent more messages per channel[6] ($p < .001$), received more messages ($p < .001$), and rated his satisfaction higher ($p < .001$) than did the peripheral *S*.

Statistically reliable differences were also found between the different support-opposition conditions for frequency of change ($p < .05$), magnitude of change ($p < .001$), message units transmitted per channel ($p < .05$), messages received ($p = .05$), and ratings of satisfaction ($p = .05$). None of the interactions was significant.

Discussion. At first glance the finding that the central *S* changed more than did the pe-

[6] Since time was constant for all trials and all *S*s, it was believed that mean number of message units per channel would reflect only differences due to the organizational aspect of positions within the net.

ripheral *S* seems to contradict the results obtained in Condition II of the second experiment; both the central *S* and the peripheral *S* who agreed with him changed *less* than did the other group members. When the differences in experimental procedure are remembered, however, the contradiction appears to be resolved. In Experiment II all *S*s were naive and each actively tried to get others to agree with her estimate, whereas in Experiment III the one naive *S* actively tried to get the confederates to agree but the confederates responded passively. That is, the confederates refused to change, but they also made no attempt to get the naive *S* to agree with their estimate.[7] In other words, it appears that the central *S* more than the peripheral *S* tries to get others to agree with his estimate, as indicated by the finding that the central *S* sent more messages *per channel* than did the peripheral *S* under similar conditions of support-opposition. If he fails, however, he changes more than does the peripheral *S*, presumably due to the greater amount of pressure exerted by the nonagreers as indicated by the fact that the central *S* also *receives* more messages than does the peripheral *S*. Failure to find a similar difference between the central *S* in Condition I and the peripheral *S* in Condition III of the second experiment appears to be due to the fact that the central *S* was sometimes successful in getting others to agree with his own estimate. When this happened, the central *S* of course did not change his estimate and differences in frequency of change due to differential amounts of pressure were cancelled out.

It is interesting to note that the results of the first two experiments agree with the results obtained by Goldberg (9) in an experimental situation which also permitted change of opinion by all *S*s. He reported that group members with less distance (the more central members) were influenced less as the group arrived at a decision.

Differences in change of estimate due to the support-opposition variable are in general agreement with those found in Experiments I and II and with the results reported of other investigations (1). Comparing the results of Experiments II and III it appears that having one supporter strengthens the resistance of an *S* relatively more than the mere reduction of the size of the opposition from three to two persons. It is further suggested that having a supporter strengthens the tendency to resist pressure relatively more than does the simple fact that the other group members are not unanimous in their opposition (cf. 1).

The ratings of satisfaction agree with the results of the first two experiments and further support the hypothesis that satisfaction is jointly determined by position within the net and the amount of support received from other group members. However, in the central position, differences in ratings of satisfaction due to differences in amount of support were not great. Apparently, the level of satisfaction was near maximum[8] in the central position even with no support from other group members.

SUMMARY

Three experiments dealing with the general problem of decision making in communication nets are reported. In the first experiment, groups of four persons were required to solve two human relations problems in three different communication nets: the star, the slash, and the comcon. The results were as follows: (*a*) *S*s in the star required significantly more time to reach a decision than did the slash, which required more time than did the comcon; (*b*) there were no significant differences among nets in ratings of satisfaction nor in number of messages transmitted, although significantly more messages were directed toward deviators than toward nondeviators; (*c*) although frequencies were too small for statistical evaluation, there were more deviators in the star than in the slash, and more in the slash than in the comcon. Also, there were more deviators in peripheral positions than in more central ones.

In Experiment II we attempted to discover whether a central *S* would change his judgment of a relatively nonambiguous stimulus more readily than a peripheral *S* when faced with different judgments by other group members. Twenty-four groups of four were placed in the star communication net and required to estimate the number of clicks in a series after discussing it with others in the group. Central and peripheral *S*s did not

[7] The latter type of situation is similar to that labeled "informational social influence" by Deutsch and Gerard (4).

[8] The highest rating possible was nine.

differ in frequency of change when all others disagreed, but when the group was equally divided both the central *S* and the peripheral *S* who agreed with him changed less frequently than did other peripheral *S*s.

The third experiment attempted to isolate the effects of the support-opposition variable upon judgmental change in different positions in the star. Sixty groups composed of one *S* and three confederates (ostensibly *S*s) were run. The task was the same as in the second experiment. Confederates either agreed or disagreed with *S*'s estimate, or made no estimate. The central *S* changed significantly more than did the peripheral *S* and both changed significantly more with increasing opposition.

From the results of these experiments the following conclusions were drawn: (*a*) The central *S* more than the peripheral *S* tries to change the opinion of those who disagree, but if he fails, he himself changes more, presumably because of his greater vulnerability to direct pressure; (*b*) the presence of one supporter strengthens the resistance of a S relatively more than the mere reduction of size of opposition *and* more than the simple fact that the opposition is not unanimous; (*c*) member satisfaction is a joint function of centrality and amount of support by other group members.

REFERENCES

1. ASCH, S. E. *Social psychology*. New York: Prentice-Hall, 1952. Pp. 450–501.
2. CHRISTIE, L. S., LUCE, R. D., & MACY, J., JR. *Communication and learning in task-oriented groups*. Cambridge, Mass.: Research Laboratory of Electronics, 1952.
3. DEUTSCH, M. An experimental study of the effects of co-operation and competition upon group process. *Hum. Relat.*, 1949, **2**, 199–231.
4. DEUTSCH, M., & GERARD, H. B. A study of normative and informational social influences upon individual judgment. *J. abnorm. soc. Psychol.*, 1955, **51**, 629–636.
5. FESTINGER, L., & THIBAUT, J. Interpersonal communication in small groups. *J. abnorm. soc. Psychol.*, 1951, **46**, 92–99.
6. FESTINGER, L., & HUTTE, H. A. An experimental investigation of the effect of unstable interpersonal relations in a group. *J. abnorm. soc. Psychol.*, 1954, **49**, 513–522.
7. GARNER, W. R. The accuracy of counting repeated short tones. *J. exp. Psychol.*, 1951, **41**, 310–316.
8. GILCHRIST, J. C., SHAW, M. E., & WALKER, L. C. Some effects of unequal distribution of information in a wheel group structure. *J. abnorm. soc. Psychol.*, 1954, **49**, 554–556.
9. GOLDBERG, S. C. Influence and leadership as a function of group structure. *J. abnorm. soc. Psychol.*, 1955, **51**, 119–122.
10. LEAVITT, H. J. Some effects of certain communication patterns on group performance. *J. abnorm. soc. Psychol.*, 1951, **46**, 38–50.
11. SCHACHTER, S. Deviation, rejection, and communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 190–207.
12. SHAW, M. E. Some effects of unequal distribution of information upon group performance in various communication nets. *J. abnorm. soc. Psychol.*, 1954, **49**, 547–553.
13. SHAW, M. E. Group structure and the behavior of individuals in small groups. *J. Psychol.*, 1954, **38**, 139–149.
14. SHAW, M. E. Some effects of problem complexity upon problem solution efficiency in different communication nets. *J. exp. Psychol.*, 1954, **48**, 211–217.
15. SHAW, M. E. A comparison of two types of leadership in various communication nets. *J. abnorm. soc. Psychol.*, 1955, **50**, 127–134.

Received May 31, 1956.

A CROSS-CULTURAL STUDY OF PERCEPTUAL PREDOMINANCE IN BINOCULAR RIVALRY[1]

JAMES W. BAGBY

Roosevelt Hospital, New York, New York

ONLY recently has research on binocular rivalry and fusion begun to explore the influence that objects which possess particular significance or meaning for an observer have upon him. Most former work had used abstract figures, circles, squares, colored patches, discrepant lines and the like. Consequently the principles of binocular fusion were elaborated on the basis of studies employing such abstract forms. A recent series of investigations by Engel (3), which employed photographs of people and objects, has demonstrated that there are important empirical and theoretical consequences when "meaningful" content is introduced in fusion and rivalry experimentation. This work strongly suggests that meaning for the perceiver is influential in the binocular resolutions achieved.

Binocular presentation of materials of disparate content and of varying subjective significance thus appears to be a technique of considerable value for psychology (1, 2, 4). The binocular conflict method offers an approach for ascertaining the role of a variety of factors in the perceptual processes. The nature of the specific resolution of two affectively charged pictures which Engel (3) presented to his subjects, for example, appeared to derive from an unconscious and spontaneous choice on the part of the subject. Although a large number of resolutions were possible, the perceptual choices were shown to reflect the subject's actual feelings and dispositions. There is need, however, for much further work to clarify the phenomena encountered in such situations of binocular conflict.

The present investigation is concerned with discovering whether the cultural characteristics of conflicting visual presentations are differentially perceived by members of different societies. Presented with a situation of simultaneous binocular rivalry, do subjects perceive more readily and consistently visual presentations of content drawn from their own culture than presentations of similar content from another society? The present study investigated this question with respect to Mexicans and Americans.

METHOD

Subjects

There were two experimental groups, one composed of 12 Mexican *Ss*, and the other composed of 12 American *Ss*. Half of each experimental group were males and the other half females. They ranged in age from 16 to 42 years. All of the *Ss* had 20-20 vision. Each *S* in the Mexican group was matched with an American counterpart as to sex, age, education, occupation and socioeconomic status. Table 1 presents some characteristics of the matched *Ss*.

An examination of Table 1 reveals that the Mexican *Ss* were largely middle-class residents of interior provincial centers and that they possessed superior educational backgrounds. The American *Ss* were residents from Northeast, Middle Atlantic, and Midwestern states with similar educational and socioeconomic backgrounds. With the exception of one matched pair, no *S* had traveled outside his own country. Thus, for both Mexican and American *Ss*, the knowledge of the opposite culture was limited to books, the mass communication media, secondhand experience, and other similar sources.

Stimulus Material

Ten pairs of photographic slides were presented. In each pair, one photograph was of a typical Mexican scene including one or more persons, and the other was a photograph of a similar American scene. The picture had been reduced or enlarged to occupy a standard 2″ by 2″ area. An attempt was made to attain similarity in the form, contour of major mass, texture, definition, and light and shadow of the paired scenes.

The following were the 10 stereogram slide pairs:

I. An American business man and a Mexican peasant.
II. A blonde American girl and a dark Mexican girl.
III. An old farm woman and an Indian peasant woman.
IV. An old farm man and an old Mexican peon.
V. An American miner and a Latin American miner.
VI. An American wedding scene of a bride and groom and a traditional Mexican wedding scene.
VII. An American farm couple and an Indian couple.
VIII. A blonde American mother with children and a dark Indian mother with children.
IX. A baseball scene and a bullfight scene.

[1] This research was made possible through a Rockefeller Foundation Grant to Princeton University for perceptual research. The writer, a research consultant to the Department of Psychology, conducted this project in Mexico during the summer of 1955.

TABLE 1
SOME CHARACTERISTICS OF MEXICAN AND AMERICAN *Ss*

Male *Ss*	Age	Education†	Occupation	SES‡	Geographic Residence
A	32	17	Statistician	Middle	Chihuahua
A′	30	16	Accountant	Middle	New York
B	19	13	Student	Middle	Durango
B′	18	13	Student	Middle	Connecticut
C	20	15	English Student	Upper	Coahuila
C′	20	15	English Student	Upper	Wisconsin
D	19	13	Mathematics Student	Middle	Jalisco
D′	18	13	Mathematics Student	Middle	Pennsylvania
E	20	14	Education Student	Middle	Morelia
E′	20	14	Education Student	Middle	New York
F	23	12	Fine Artist	Upper	Mexico City
F′	22	13	Commercial Artist	Upper	Massachusetts
Female *Ss*					
G	27	17	College Teacher	Upper	Torreon
G′	27	16	High School Teacher	Middle	Kentucky
H	42	16	Language Teacher	Middle	Coahuila
H′	40	17	Language Teacher	Middle	New England
I	33	16	Elementary Teacher	Middle	Chihuahua
I′	32	16	Elementary Teacher	Middle	Illinois
J	16	11	Student	Upper	Guadalajara
J′	17	12	Student	Upper	New Jersey
K	27	12	Artist	Middle	Jalisco
K′	30	16	Art Teacher	Middle	Iowa
L	29	16	Teacher	Middle	Guanajuato
L′	31	16	Teacher	Middle	Colorado

* *Ss* matched: A, Mexican—A′, American; B, Mexican—B′, American, etc.

† Numeral refers to years of formal education. Thus, 11 is equivalent to junior year in high school, 17 is equivalent to one year beyond college graduation, etc.

‡ SES refers to socioeconomic status in three groupings: upper, middle and lower classes.

Reprinted by special permission of the SATURDAY EVENING POST. *Copyright c 1955 by the Curtis Publishing Company.*

Reprinted by special permission of Alfred A. Knopf. Copyright by Manuelo Gonzales, photographer.

FIG. 1. EXAMPLE OF STIMULUS PAIR: STEREOGRAM IX

X. A young American boy and a Mexican boy. Pair IX is reproduced in Fig. 1.

Apparatus

The stereogram slide pairs were presented to *S* by means of a prism lens stereoscope enclosed in a light-tight box, like that more fully described by Bagby and Engel (2, 4). Each eye was independently presented with but one photograph of a pair. The slide holder was adjustable to permit the proper setting for a normal binocular fusion of the two stimuli. The intensity of illumination upon either slide could be independently varied, but for the present set of data it was equated.

Procedure

The *S* was informed that he was to view a series of slides in the apparatus and was to tell *E* what he saw. The *S* was cautioned to look with both eyes and to keep

his forehead pressed against the metal view piece. When the preliminary questions regarding procedure were answered to *S*'s satisfaction he was placed in viewing position.

The initial presentation was stereoscopic slide 33 of the Titchener series (7). The *S* was asked to adjust the setting so the white circle (in one visual field) was centered as nearly as possible in the black circle (presented to the other field). When this had been achieved a slide from the series of Engel's "male faces" (3) was presented and *S* was asked what he saw. All described a single male face. The *S* was then asked if the face appeared in clear focus. Ordinarily there was some minor adjustment to be made in the setting of the apparatus to obtain proper fusion.

At this point *S* was told that the apparatus was adjusted for his own vision and that the experiment *per se* was to begin. The sequence of Mexican-American stereogram slides was then presented, each being exposed for a 60-sec. viewing period. To control for eye dominance, the left-eye-right-eye positions of the Mexican and American scenes were randomized within the series. In terms of *S*'s running descriptive report of what he saw, *E* recorded the relative predominance of the right or left stimulus object. The basic data, however, derive from the initial view or first 15-sec. exposure of the slides.

RESULTS AND DISCUSSION

A variety of reports was obtained from *S*s:

1. In the simplest (and relatively infrequent) case *S* reported seeing but one slide of a pair, and solely described its features. This slide could readily be classified as "predominant" or "preferred" by *S*.

2. In other cases one slide of the pair would be described very thoroughly, after which *S* would begin to say he saw "something else," gradually describing the features of the other slide while reporting that each picture "comes and goes." In the case of such "rivalry," the preference or predominance of one slide was determined by: (a) the picture first reported, and (b) *S*'s statements as to which picture seemed to be present most of the time.

3. There were also reports of a single picture which was an "admixture" of the two pictures presented. "Admixture" reports were rare for the initial view; usually they came during total exposure after a transition from marked predominance. It was not difficult to determine the preferred slide since *S*'s report of content was definitive.

4. There were also reports of the two pictures being seen at once with one superimposed on the other. This condition was reported for first view on only three occasions.

Table 2 gives the over-all results. The Wilcoxon (8) test was used for determining the significance of differences between the selections made by the Mexican and American *S*s. Significance was at the .01 level of confidence.

TABLE 2

DIFFERENTIAL SELECTION OF AMERICAN CONTENT IN TEN STEREOGRAMS

Nationality	Number of *S*s Selecting American Content in given Number of Stereograms				
	8-10	5-7	2-4	0-1	Total
Mexican *S*s	0	0	5	7	12
American *S*s	9	3	0	0	12
Total	9	3	5	7	24

TABLE 3

OVER-ALL PERCEPTUAL PREDOMINANCE IN THE TEN STEREOGRAM PAIRS

	No. Where Mexicans Dominate	No. Where Americans Dominate	Total (6×10)
Mexican males (6)	44	16	60
Mexican females (6)	45	15	60
American males (6)	7	53	60
American females (6)	12	48	60

Since the initial time interval was rather short, it may be suggested that culturally familiar subject matter was preferentially reported in cases 3 and 4 without the selection being strictly perceptual. This possibility was checked by excluding these cases from the analysis; the over-all results remain unchanged with significance at the .02 level.

An additional check on the preferential reports for cases 3 and 4 was conducted on all *S*s using the same stimulus materials. This consisted of re-exposing the materials and varying the intensity of light on the two pictures (2, pp. 10–11). The degree to which illumination of the less preferred picture had to be increased to offset the advantage of the other provided a quantitative measure of perceptual predominance. The direction only, and not the relative strength of the indexes, was used in further support of the predominance report derived from the first-view-in-time. In no instance were the original criteria for determining preference, more especially cases 3 and 4, contra-indicated.

Table 3 gives the results from the intensity variation.

Since the probability of equal perceptual predominance was rejected at the .01 level for both *S*s and content, the conclusion seems

justified that *S*s report scenes of their own culture as predominant in binocular rivalry over scenes from another culture. The national cultural differences appear critical in affecting perceptual predominance in the majority of the stereogram slide pairs. There were only three pairs in which cultural preference was not marked (i.e., I, V, and X). One can only speculate concerning the less marked cultural influence in the latter pairs, which could involve the physical characteristics of the photographs (e.g., greater clarity or better definition of the prints), or possibly greater "appeal" of the particular individuals depicted regardless of cultural setting. However, the major finding is clear.

Various theoretical interpretations are possible for describing the findings of the present investigation. The position here favored, however, is that advanced by Ames, Cantril, and the transactional school who regard perception as being fundamentally determined by previous, rather than present, experience (5, 6). In transactional perceptual theory the role of meaning is accorded a central position in the perceptual processes. Differences in ways of perceiving come about as a consequence of differences in past experiences and purposes. These in turn emerge from influences in the home, in the school, and in the various groups with which an individual identifies. Thus, under conditions of perceptual conflict as found in the binocular rivalry situation, those impingements possessing the more immediate first-person meaning would be expected to predominate in visual awareness. The findings of the present experiment seem accountable in these terms. While greater familiarity with the objects and scenes from the *S*'s own culture provided a more dominant set for perceiving those materials in preference to the relatively less familiar, other variables such as mood, need, and self-reference values undoubtedly entered into the actual selections. The best explanation of the present data, therefore, would seem to be in terms of past experience. The accumulated past experience of the individual within the characteristic settings and with typical individuals of his own culture should make pictorial scenes from his own culture possess greater personal significance at the visual level than that possessed by the less familiar situations and individuals of another society.

SUMMARY

Twelve Mexican and twelve matched American subjects were simultaneously presented with a series of ten stereogram slide pairs of similar scenes, one from Mexico and one from the United States. Under these conditions of experimentally induced binocular rivalry, it was found that scenes from the *S*'s own culture tended to be perceptually predominant. The findings were accounted for in terms of the demonstrable role of personal significance in perceptual processes.

REFERENCES

1. BAGBY, J. W. The relative roles of information and action in the genesis of a perception. Doctor's dissertation, Columbia Univer., 1955.
2. BAGBY, J. W. A perceptual study of cross-cultural attitudes in Mexico and the United States. Advance report, Psychol. Dept., Princeton Univer., March 1956.
3. ENGEL, E. Binocular conflict and resolution. Doctor's dissertation, Princeton Univer. 1955.
4. ENGEL, E. The role of content in binocular resolution. *Amer. J. Psychol.*, 1956, **69**, 87–91.
5. ITTELSON, W., & CANTRIL, H. *Perception: a transactional approach*. New York: Doubleday, 1954.
6. KILPATRICK, F. P. Human behavior from the transactional point of view. Hanover, N. H.: Institute for Associated Research, 1952.
7. Titchener Series. *Stereoscopic slides*. Chicago: C. H. Stoelting Company, 1942.
8. WILCOXON, F. *Some rapid approximate statistical procedures*. New York: American Cyanamid Company, 1949.

Received June 8, 1956.

PSYCHOGENIC DEAFNESS AND PERCEPTUAL DEFENSE[1]

MILTON L. KLEINMAN[2]

New Jersey State Diagnostic Center, Menlo Park

FROM investigations of the relationship between motivation and perception has come the concept of perceptual defense (1, 8, 10), an adaptive process postulated to account for the raised perceptual thresholds obtained when subjects are confronted with unfavored or anxiety evoking stimuli. There has been an attempt, however, to explain these increased thresholds on the basis of the relative frequency of the stimulus words used, and on the reluctance of subjects to repeat aloud stimulus words that may be considered socially taboo (3, 4, 14).

The aim of the present study is to investigate whether characteristic defensive reactions to anxiety evoking stimuli may be demonstrated through changes in auditory perceptual thresholds. It is hypothesized (*a*) that perceptual defense against anxiety producing, verbal stimuli can serve to elevate auditory thresholds, and (*b*) that individuals whose auditory thresholds are characteristically raised in their daily life as a defense against anxiety provoking stimuli can be identified on the basis of their perceptual behavior in the laboratory.

In order to test these hypotheses the following experimental criteria must be satisfied:

1. The experimental subjects must have similar modes of reacting to inimical auditory stimuli.

2. All stimulus words used must be a part of socially approved everyday speech so that there can not be any inhibition against their use in the experimental situation.

3. The incidence of occurrence in the language of the stimulus words must be carefully controlled.

4. The auditory intelligibility of the stimulus words must be equated so far as possible.[3]

METHOD

To satisfy the first criterion the experimental group consisted of 22 *Ss* with the diagnosis of psychogenic deafness, whose auditory acuity fluctuated without any observable organic basis. It is widely believed that this condition occurs when hearing becomes involved in a basic emotional conflict that the patient desires to repress (7, 11, 15). In terms of personality structure, psychogenic patients have been described as immature, dependent persons who have difficulty with expressing themselves in important interpersonal relationships (12). Although highly dependent upon authority figures, they are unable to recognize this dependency.

The control group consisted of 22 *Ss* with organic hearing loss who showed no inconsistencies in their ability to hear. All *Ss* were male veterans of World War II who were patients of the Audiology Clinic at the New York Regional Office of the Veterans Administration. The differential diagnosis was made after a thorough evaluation by the clinic staff which consisted of otologists, clinical audiologists, teachers of speech, teachers of speech (lip) reading, a psychiatrist, a social worker, and a clinical psychologist.

In order to obtain socially approved stimulus words, current literature was searched for a passage of potential interest to World War II veterans involving a sharp conflict with authority. Since conflict with severe authority in relation to deep dependency needs is thought to characterize the psychogenic patients, such a passage might be expected to arouse their anxiety. A passage meeting these specifications was found in the novel by Norman Mailer, *The Naked and the Dead*. The scene in which the homosexually inclined General orders his aide to pick up his cigarette was first adapted for dramatic presentation[4] and then recorded by professional actors. A count of the important words revealed eleven occurring one or more times. To these were added the words "angry," "force," and "father," which were considered easy associations to the playlet by three clinical psychologists who listened to the final recorded version. These fourteen words formed the crucial or anxiety evoking list (Table 1). An equal number of unrelated words matched for rhythm of syllabication, accent, and incidence of occurrence in the language according to the Thorndike-Lorge list (14) formed the neutral list.

Since two words reaching the ears at the same level of sound intensity are not necessarily equal in intelligibility, a correction was needed. In a preliminary experiment an intelligibility score, designed to produce homogeneous intelligibility was obtained for each of the

[1] An abridgment of a dissertation in the Department of Psychology, submitted to the faculty of the Graduate School of Arts and Science, New York University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy. Degree conferred in June 1953.

[2] The author wishes to express his sincere appreciation to Professor Leland W. Crafts for his guidance throughout the course of this investigation. He is also indebted to Dr. Moe Bergman who was especially helpful in preparing the recordings. This research was conducted at the Audiology Clinic, New York Regional Office of the Veterans Administration.

[3] The bases for the development of these criteria can be found in the original dissertation, p. 9 f.

[4] This text appears on p. 36 f. of the dissertation.

TABLE 1
LIST OF CRUCIAL AND NEUTRAL WORDS

Crucial	f*	I.S.†	Neutral	f*	I.S.†
Robert	49	9	Joseph	58	6
power	AA	3	dollar	AA	6
resistance	26	0	delightful	32	3
superior	A	0	immediate	A	6
cigarette	22	3	chocolate	20	0
obedience	15	0	recovery	16	3
unfair	8	0	rainfall	7	3
inferior	19	0	phenomenon	18	3
prison	A	12	delight	A	3
ashtray	37 17	6	doormat	AA 18	6
lecture	28	0	debate	24	6
angry	A	0	empty	A	6
force	AA	18	chair	AA	0
father	AA	3	daughter	AA	3

* Thorndike-Lorge frequency count per million (A = over 50, AA = over 100).

† Intelligibility score expressed in decibels of attenuation needed to produce homogeneous intelligibility.

stimulus words by a modification of the method described by Harris (2). These are also shown in Table 1.

To prepare the stimulus words for presentation to the *S*s, they were first recorded on a disc in random order, by a trained speaker who monitored his voice with a sound level meter. Each word was preceded by the carrier phrase, "The next word is . . . ." Each of these words was then re-recorded from the original ten times with a five-second interval between each repetition. During this re-recording, slight variations in peak intensity level due to human errors in monitoring were overcome electronically. The words were then re-recorded again with an attenuator in the circuit to include the correction for intelligibility. To eliminate relative radial distortion and surface wear, the final re-recordings were made on tape. The stimulus words were arranged in two different random orders, one for the initial series of threshold determinations, and one for the final series. Since they all were re-recorded from the same monitored original disc, each of the stimulus words and its repetitions was identical.

Apparatus

The test chamber was a soundproof room with an ambient noise level of less than 20 db. Facing the *S* was an Altec-Lansing 604B loud speaker, in a base reflex baffle. There was also a Shure model 556 cardiod microphone to pick up his responses. The *S* could be observed from the control room through a series of observation windows. This adjoining control room was sound treated, and contained a Jensen 8-in. monitor speaker through which the *E* could hear *S*'s responses. It also contained the high fidelity sound reproducing equipment, which included a Gates CB7 16-in. transcription turntable with an Audak pickup and transcription arm for playing the recorded playlet, and a Magnecorder recording and playback unit type PT6-AM with a type PT6-J amplifier for playing the tape. The turntable fed to a Langevin preamplifier, which with the tape recorder fed to a Daven attenuator calibrated in 3 db steps. These, in turn, fed through a Brook 2A3 power amplifier to a Daven output attenuator calibrated in 10 db steps, which fed to the loud speaker in the test chamber. The range of sound intensity extended from −10 db to +98 db, and could be controlled to within a difference of one decibel.

Procedure

Each *S* sat in the test chamber and faced the *E*, who sat at the control console in the control room. The *S*'s speech-reception threshold was first obtained by the standard method used in audiology clinics (5). This afforded *S* an opportunity to become accustomed to the test chamber, provided practice in listening through the equipment, and provided a reference point for the experiment itself. Recorded instructions, to say each word that was heard, were followed by one practice trial. The threshold for each stimulus word was determined by the method of limits (ascending series). Each word was first presented at 9 db below the *S*'s threshold. For every repetition of the word, the intensity was increased 3 db, until *S* reported it correctly once.

After the recognition thresholds for the twenty-eight stimulus words of the initial series were obtained, *S* was told he was going to hear an excerpt from a broadcast. A transcription of the playlet was then presented at an intensity level sufficiently above his threshold to be comfortably heard, despite any hearing loss. Upon completion of the recorded playlet, he again heard recorded instructions, and one practice trial preceded the final series of threshold determinations.

RESULTS

The organic and psychogenic *S*s were closely alike with respect to age, education, and speech-reception threshold. Mean differences were all insignificant.

The change in threshold for the crucial words and for the neutral words was computed for each *S* by subtracting the combined threshold on the final series from the combined threshold on the initial series for each set of words. Use of the combined thresholds yielded the total change in threshold for each *S*, for both word types, and avoided fractions in later computations.

For each of the organic *S*s, the combined threshold for both sets of words, crucial and neutral, was higher on the initial series than on the final series. For the psychogenic *S*s, the crucial words yielded higher combined thresholds for the initial than for the final series in 17 cases. The remaining 5 *S*s had higher combined thresholds on the final series than on the initial series. The neutral words yielded higher combined thresholds for the initial than for the final series for 21 of the 22 *S*s.

Group Reductions in Threshold

Organic group. The mean change in threshold for the organic group between the initial and the final series was 89.59 for the crucial words. For the neutral words this was 78.27. The difference between the means was 11.32, which was not significant The reduction in threshold for the crucial words was greater than the reduction in threshold for the neutral words.

Psychogenic group. The mean change in threshold for the psychogenic group between the initial and the final series was 35.32 for the crucial words. For the neutral words this was 73.23. The difference between the means was −37.91 with $p < .07$. The reduction in threshold for the neutral words was greater than the reduction in threshold for the crucial words.

Group Differences in Mean Threshold Reduction

Comparing the mean differential reduction in threshold for the crucial vs. neutral words for the organic group with the corresponding value for the psychogenic group, a higher-order difference of 49.23 is obtained. The critical ratio of this difference was 6.835, which was significant at the .01 level. A test for the homogeneity of variance was not significant. The difference between the performance of the two groups, therefore, is due entirely to the difference between the means.

TABLE 2

COMPARISON OF THE DIFFERENCES BETWEEN TYPES OF WORDS AND BETWEEN GROUPS

Type of Hearing Loss	Difference Between Reductions in Threshold for Crucial and Neutral Words			t	(f)
	Mean Reduction	*SD*	Diff.		
Organic*	11.32	19.9	49.23	6.835†	(1.728)
Psychogenic*	−37.91	26.1			

* $N = 22$.
† $p < .01$.

DISCUSSION

Since none of the crucial or neutral words used in this experiment could be considered taboo or unfit for polite usage, the reluctance to respond that Howes & Solomon (3) found in McGinnies' (8) experiment was eliminated. Also, because each word was not only identical for the initial and final series, but was also identical for every repetition of itself, any influence of variability among the stimulus words themselves, such as occurred in an earlier experiment of Vanderplas and Blake (16) was also eliminated.

Although there was the opportunity for a practice effect with both sets of words, there was a greater opportunity with the crucial words because of their inclusion in the intervening playlet. The over-all practice effect was significantly greater for the organic than for the psychogenic *Ss*. In addition, 68 per cent of the organic *Ss* ($p < .05$) had a greater reduction in threshold for the crucial than for the neutral words. On the other hand, the psychogenic *Ss* had a greater mean reduction in threshold for the neutral than for the crucial words.

The differences between the mean reductions in threshold were statistically significant for six of the psychogenic *Ss*. In terms of response frequency, all the psychogenic *Ss* showed a greater reduction in threshold for the neutral than for the crucial words. This negative effect of the increased opportunity for practice with the crucial words may be due to the relationship between the content of the play and deep unresolved conflicts.

The first hypothesis, that perceptual defense against anxiety producing, verbal stimuli can serve to elevate auditory thresholds, is not fully supported. Under the conditions of this experiment a tendency towards a decreased practice effect was found instead. This result was similar to data reported by Postman and Bruner (9) in their study of visual perception under stress. They also found a decreased practice effect for the experimental group, compared with the control group, following the stress situation. The present study was unlike that of Postman and Bruner, however, in that in their experiment the effect of the stress situation was general, while in the present experiment the anxiety was restricted to specific words. In this sense the present experiment resembles more closely the study of McGinnies (8), where seven emotional words were dispersed among eleven neutral words, with results consistent with those reported here.

The significant differential performance of the two groups supports the second hypothesis, that individuals whose auditory thresholds are characteristically raised as a defense against anxiety provoking stimuli can be identified on

the basis of their perceptual behavior. This finding is consistent with perceptual defense theory. The decreased reduction in threshold for the crucial words by the psychogenic *Ss* may be related to a perceptually manifested defense against the deeper conflicts to which the crucial words were related.

By a somewhat different technique, Lazarus, Eriksen, and Fonda (6) found that the accuracy of auditory recognition was related to the clinical evaluation of dynamic personality processes. The present results are consistent with these findings. In the present experiment, the objective characteristics of the perceptual stimuli were held constant. The major variable was the inferred characteristic difference in dynamic personality processes of the two groups of *Ss*. The significant difference between the groups may therefore be attributed to the influence of dynamic personality processes upon auditory perceptual thresholds.

SUMMARY

The purpose of this experiment was to investigate whether perceptual defense against anxiety evoking, verbal stimuli is observable through a change in auditory perceptual thresholds. An experimental group of 22 *Ss* with psychogenic hearing loss, and a matched control group with organic hearing loss were employed. Crucial and neutral words, matched for rhythm of syllabication, accent, frequency of occurrence in the language and corrected for homogeneous intelligibility, were presented in two random orders. The crucial words were centrally involved in the text of a playlet concerned with a sharp conflict with authority, which was presented by transcription after the auditory thresholds for the stimulus words had been obtained. After *Ss* had heard the playlet the threshold for each of the stimulus words was again obtained.

The difference between the psychogenic group and the organic group with respect to the differential mean reduction in threshold for the crucial and neutral words was significant ($p < .01$). The results are interpreted as being generally consistent with perceptual defense theory. The greater perceptual difficulty evidenced by the psychogenic *Ss* with the crucial words was considered a perceptually manifested defense against deep conflicts to which these words had been related by the playlet.

REFERENCES

1. ERIKSEN, C. W. Perceptual defense as a function of unacceptable needs. *J. abnorm. soc. Psychol.* 1951, **46**, 557–564.
2. HARRIS, J. A. Some suggestions for speech reception testing. *Arch. Otolaryng.*, Chicago, 1949, **50**, 388–404.
3. HOWES, D. H., & SOLOMON, R. L. A note on McGinnies' emotionality and perceptual defense. *Psychol. Rev.*, 1950, **57**, 229–234.
4. HOWES, D. H., & SOLOMON, R. L. Visual duration threshold as a function of word-probability. *J. exp. Psychol.*, 1951, **41**, 401–410.
5. HUDGINS, C. U., HAWKINS, J. C., KARLIN, J. E., & STEVENS, S. S. The development of recorded auditory tests for measuring hearing loss for speech. *Laryngoscope*, 1947, **57**, 57–89.
6. LAZARUS, R. S., ERIKSEN, C. W., & FONDA, C. P. Personality dynamics and auditory perceptual recognition. *J. Pers.*, 1951, **19**, 471–482.
7. MARTIN, N. A. Psychogenic deafness. *Ann. Otol., etc.*, St. Louis, 1946, **55**, 81–87.
8. MCGINNIES, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244–251.
9. POSTMAN, L., & BRUNER, J. S. Perception under stress. *Psychol. Rev.*, 1948, **55**, 314–323.
10. POSTMAN, L., BRUNER, J. S., & MCGINNIES, E. M. Personal values as selective factors in perception. *J. abnorm. soc. Psychol.*, 1948, **43**, 142–154.
11. RAMSDELL, D. A. The psychology of the hard of hearing and the deafened adult. In H. Davis, (Ed.), *Hearing and deafness*. New York: McGraw-Hill, 1947.
12. SLOTE, W. H. The personality of the psychogenic hard of hearing adult. Unpublished doctor's dissertation, New York Univer., 1951.
13. SOLOMON, R. L., & HOWES, D. H. Word frequency, personal values and visual duration thresholds. *Psychol. Rev.*, 1951, **58**, 256–270.
14. THORNDIKE, E. L., & LORGE, I. *The teachers word book of 30,000 words*. New York: Teachers College, Columbia Univer., 1944.
15. TRUEX, E. H., JR. Psychogenic deafness. *Conn. Med. J.*, 1946, **10**, 907–915.
16. VANDERPLAS, J. M., & BLAKE, R. R. Selective sensitization in auditory perception. *J. Pers.*, 1949, **18**, 252–266.

Received June 11, 1956.

EXPECTANCY CHANGES IN SKILL AND CHANCE SITUATIONS[1]

E. JERRY PHARES

Kansas State College

ANY theory of personality employing a construct of expectancy must be prepared to state the conditions under which expectancies change. Obviously, one such condition is the situation, or more precisely, the individual's categorization of the situation. Recently, considerable attention has been given to an analysis of situational variables as they affect psychological testing and behavior generally (3, 5, 8, 9, 11, 13, 15). This paper explores the relationship between expectancy changes and situational variables, and makes a preliminary effort toward establishing a classification of situations in terms of their effect on expectancies.

The study was stimulated by Rotter's theoretical formulation of the nature of social learning (12), in which human behavior is conceived as being a product of choices. Which choice and thus which behavior occurs is for Rotter a function of both the value of the reinforcement and the expectancy that a given behavior will be successful for a given situation or class of situations. Situations, experimental or natural, are composed of cues that lead to expectancies for various behavior-reinforcement sequences. The individual's categorization of the situation determines his expectancy that a given behavior will be followed by a given reinforcement. How the individual categorizes a situation depends upon what prior situations he sees as having been the same, the amount of his prior experience in these situations, and his overall, generalized expectancy.

Lasko (7) recognized a distinction between learning in situations where the effects following a behavior are a function of the behavior itself, and learning in situations where the subsequent effect or reinforcement is controlled by someone else according to a prearranged sequence. For purposes of this study situations may accordingly be defined as involving skill or chance. A skill situation is characterized by the fact that the occurrence of a reinforcement is related by the subject to his performance, while a chance situation is one wherein *S* perceives the task as so difficult that being right or wrong is not so much a function of skill as it is luck or *E*'s whim. The occurrence of reinforcement, then, is beyond *S*'s control.

In the skill situation *S* should perceive his performance as providing a basis for generalization to future performance. Since he is the effective agent in the situation he may generalize from past similar experiences. If he fails on a given trial, there should be a decrement in expectancy of future success because his scores are a function of his skill. Furthermore, his experience in the situation gives him successively more adequate conceptions of what to expect on subsequent trials. Thus, the number of trials in a skill situation should be a potent factor in determining subsequent expectancies. In chance situations, on the other hand, none of these statements should hold.

For testing this formulation concerning the situational determinants of expectancy changes the following major hypothesis was framed: 1. Increments in expectancy following successes and decrements in expectancy following failures are greater in a situation generally categorized by *S* as involving his skill than in a situation generally categorized by him as chance.

Stemming from the major hypothesis were three subsidiary predictions. 2. A greater proportion of "unusual" shifts in expectancy is predicted in the chance situation than in the skill situation. "Unusual" shifts refer to increments in expectancy after failure or decrements after success. In a chance situation, *S*s may be expected often to invoke the "gambler's fallacy," basing their expectancy upon fallacious reasoning such as, "Well, I've been right three times in a row, so it's about time I was wrong." The result is a decrement in expectancy following success; a decrement based not so much upon the previous trial as upon a fallacious view of the odds for success and failure.

While the major hypothesis concerns incre-

[1] This paper is adapted from a dissertation submitted in partial fulfillment of the requirements for the Ph.D. in the Graduate School of the Ohio State University. A special debt of gratitude goes to the director of this dissertation, Dr. Julian B. Rotter.

ments after success and decrements after failure (the "usual" case), it is further predicted: 3. That increments and decrements in expectancy are greater in skill situations than in chance situations regardless of whether they are "usual" or "unusual." It is also hypothesized: 4. That the frequency or number of shifts is greater in the skill situation.

Several studies (4, 6, 10, 16) have sought to establish particular relationships between goal value and expectancy, obtaining divergent results. Seemingly, different experimental designs result in different categorizations of the situation by *S*s, which lead, in turn, to different results. A situational analysis in terms of the chance vs. skill dichotomy seems promising as an approach to establishing order in this research area.

METHOD

Seventy-seven female students from introductory psychology courses were used as *S*s. Two experimental tasks were devised so that each could be structured by instructions as either chance or skill.

Task A (*Color*). Ten 2-in. by 3-in. sample paint patches of varying shades of gray (referred to as the standard colors) were mounted in random order of lightness to darkness on different colored 4-in. by 6-in. sheets of construction paper, which in turn were placed on a large drawing board in two rows of five cards each. Thirteen of these same color patches reduced in size to 2 in. by 1 in. (no one shade occurring more than twice) were pasted, separately, on 3-in. by 5-in. white filing cards. Each card's patch matched exactly the color of one on the board.

Task B (*Length*). Materials for this task consisted of ten ½-in. wide strips of black construction paper ranging from 1 in. to 2 ½ in. in length (referred to as the standard lengths). The difference in length between any two adjacent size strips was ½ in. Each of the ten strips was pasted, separately, on 3-in. by 5-in. white filing cards, at varying angles. All of the cards were then mounted at differing angles on a large drawing board. Thirteen of these same strips (color other than black, and no one length occurring more than twice) were pasted at varying angles on 2-in. by 2 ½-in. white cards. Each card's strip was exactly the same length as one on the board.

Procedure. Four experimental groups were used to which *S*s were assigned in rotation according to their order of appearance. The procedure is summarized in Table 1. Both order of presentation of tasks and the manner in which each task was structured were balanced. The tasks required *S* to match the matching colors (or lengths) with the standard colors (lengths) on the boards. The skill instructions described the tasks as being difficult, but ones in which score reflected ability. The *S* was told that correct matchings were possible and that some people did well and others poorly depending on skill. The chance task, on the other hand, was described as being so difficult that it did not differentiate between people of high matching and low matching ability. The *S* was told that no one made consistently correct matchings and that success here had been found to be only a matter of luck. This basic structure was repeated in different ways several times and the distinction between tasks reiterated before matchings began.

Each *S* was read one of four sets of instructions depending upon to which experimental group she was assigned, and was given thirteen trials of matching on each task. The sequence of reinforcement was fixed, regardless of task or *S*, and is shown in Table 2.

Although reinforcement in both situations in fact bore a chance relationship to *S*'s performance, the instructions apparently exerted a strong effect. The *S*s did not recognize that they were being reinforced similarly in the two situations. Thus, because of initial structure, *S*s categorized the tasks differently and therefore gave differing interpretations to their successes or failures. Many spontaneous comments attested to the fact that none of the *S*s questioned the report of failure or success for any given trial even though in reality it may have contradicted their actual performance. Both tasks were difficult and novel enough to lend credence to *E*'s statements.

The measure of expectancy used was a betting technique developed by Castaneda (2). This method correlates .99 with verbal methods, although the betting method yields somewhat higher expectancies (14). Before each matching trial *S* was asked to bet from one to ten chips as to whether she would be correct in her matching. During the betting *E* recorded the number of chips bet each time and also a cumulative total of chips won to lend reality to the situation.

TABLE 1

EXPERIMENTAL GROUPS AND ORDER OF PRESENTATION OF TASKS

Groups	Order of Presentation	
	1st Task	2nd Task
1.	Color (chance instructions)	Length (skill instructions)
2.	Color (skill instructions)	Length (chance instructions)
3.	Length (chance instructions)	Color (skill instructions)
4.	Length (skill instructions)	Color (chance instructions)

TABLE 2

FIXED SEQUENCE OF REINFORCEMENT

	Trial												
	1	2	3	4	5	6	7	8	9	10	11	12	13
x*	x	–	x	–	–	x	x	x	–	–	–	x	

* x refers to a positive reinforcement (correct matching), success. – refers to a negative reinforcement (incorrect matching), failure.

RESULTS AND DISCUSSION

The data for the major hypothesis was analyzed in the following manner: A score was computed for each *S*, in each type of situation by taking the absolute amount of all "usual" increments and decrements in expectancy (bets) and dividing by the number of "usual" shifts. The mean magnitude of "usual" shifts in the skill situation was 1.8 while the corresponding mean magnitude in the chance situation was 1.5. When the *t* test is applied,[2] the difference between means is highly significant ($t = 2.6$, $p = .004$).

The mean proportion of "unusual" shifts to total number of shifts was .16 in the skill situation and .25 in the chance situation. The results are in the predicted direction but not statistically significant ($t = 1.5$, $p = .07$). It was also predicted that expectancy shifts would be greater in the skill than in the chance situation. Here, the amount of *all* increments and decrements were summated without regard to sign and divided by twelve for each *S*. Means of .93 and .79 were thus obtained for the skill and chance situations, respectively, the difference between them being significant in the predicted direction ($t = 2.3$, $p = .01$). Finally, the frequency of *all* shifts in expectancy was greater in the skill situation (mean frequency = 6.12) than in the chance situation (mean frequency = 5.49). The *t* here was 2.1 with *p* at the .02 level.

Only the hypothesis with respect to proportion of "unusual" shifts failed to be confirmed at an acceptable level of significance; even in this case results in the predicted direction of borderline significance were obtained. The findings thus support the view that categorizing a situation as skill leads *S* to use the results of his past performance in formulating expectancies for future performances. In chance situations, on the other hand, past performance does not provide a basis for generalization to future trials since *S* is not the effective agent in obtaining reinforcements. These results, limited to situations defined in terms of chance and skill, suggest the wider fruitfulness of a situational analysis of expectancy changes. Moreover, the need seems apparent for some sort of representative sampling of situations as urged by Brunswik (1), going beyond the narrow range investigated in the present study.

[2] One-tailed tests of the distribution of *t* were used throughout this research.

SUMMARY

This study attempted to demonstrate the effect of situational variables on expectancy changes. From Rotter's social learning theory it was hypothesized that expectancies for future reinforcement show greater changes following reinforcement in a skill situation than in a chance situation. This prediction was based on the interpretation that since obtained scores (reinforcements) in a skill situation are a result of performance, *S*s should utilize past performances as a basis for future expectancies for reinforcement in such a situation. In a chance situation, obtained scores are not controlled by the *S*, thus providing little basis for generalization to future trials.

Two experimental tasks were designed involving (*a*) the matching of colors and (*b*) the matching of lengths of lines. Seventy-seven female *S*s were used, assigned to four groups in a systematic design that varied order of task and of instructions creating skill or chance situations. Each *S* had 13 trials on each task, and all received the same pattern of positive and negative reinforcement.

The major finding was that the chance situation produced smaller expectancy changes than the skill situation. Moreover, the frequency of expectancy shifts was greater in the skill situation.

REFERENCES

1. BRUNSWIK, E. The probability point of view. In M. H. Marx (Ed.), *Psychological theory*. New York: Macmillan, 1951.
2. CASTANEDA, A. A method for measuring expectancy as conceived within Rotter's social learning theory of personality. Unpublished master's thesis, Ohio State Univer., 1951.
3. CHEIN, I. The environment as a determinant of behavior. *J. social. Psychol.*, 1954, **39**, 115–127.
4. CRANDALL, V. J., SOLOMON, D., & KELLOWAY, R. Effects of objective probability and reinforcement value on expectancy statements and decision time. Paper read at Midwest. Psychol. Ass., 1955.
5. GIBBY, R. G., MILLER, N. R., & WALTER, E. L. The examiner's influence on the Rorschach protocol. *J. consult. Psychol.*, 1953, **17**, 425–428.
6. IRWIN, F. W. Stated expectations as functions of probability and desirability of outcome. *J. Pers.*, 1953, **21**, 329–335.
7. LASKO, A. The development of expectancies under conditions of patterning and differential reinforcement. Unpublished doctor's dissertation, Ohio State Univer., 1952.

8. LORD, E. E. Experimentally induced variations in Rorschach performance. *Psychol. Monogr.*, 1950, **65,** No. 10, 1–34.
9. LUCHINS, A. S. Situational and attitudinal influences on Rorschach responses. *Amer. J. Psychiat.*, 1947, **15,** 283–291.
10. MARKS, R. W. The effect of probability, desirability, and "privilege" on the stated expectations of children. *J. Pers.*, 1951, **19,** 332–351.
11. PHARES, E. J., & ROTTER, J. B. An effect of the situation on psychological testing. *J. consult. Psychol.*, 1956, **20,** 291–293.
12. ROTTER, J. B. *Social learning theory and clinical psychology.* New York: Prentice-Hall, 1954.
13. ROTTER, J. B. The role of the psychological situation in determining the direction of human behavior. In M. R. Jones (Ed.), *Nebraska symposium on motivation*, 1955. Lincoln: Univer. of Nebraska Press, 1955.
14. ROTTER, J. B., FITZGERALD, B. J., & JOYCE, J. N. A comparison of some objective measures of expectancy. *J. abnorm. soc. Psychol.*, 1954, **49,** 111–114.
15. SARASON, S. B. *The clinical interaction.* New York: Harper, 1954.
16. WORELL, L. The effect of goal value upon expectancy. *J. abnorm. soc. Psychol.* 1956, **53,** 48–53.

Received June 15, 1956.

SEX DIFFERENCES IN SOCIAL MEMORY TASKS

SAM L. WITRYOL AND WALTER A. KAESS
University of Connecticut

IN AN earlier study (2) the present investigators found marked sex differences favoring females in performance upon two social recall tasks: the memory for names-and-faces subtest of the George Washington University Social Intelligence Test (5), and a names-and-faces miniature life situation. The purpose of the present investigation was to analyze sex differences on three names-and-faces memory tasks: (*a*) the GWU subtest, (*b*) the miniature life situation, and (*c*) a new memory-for-names-and-faces picture test developed by the writers. In the previous research (2) the first two tasks had yielded results reflecting small variances and insufficient ceilings. The third instrument was developed to provide greater discrimination by increasing the number of items, including pictures of both sexes, and minimizing spatial cues through altering the picture pose from memory task to subsequent recall test. In contrast with the GWU subtest which requires the correct association of one of a series of *pictures* with each name, this new test requires the subject to associate one of a series of *names* with each picture-portrait. A recognition as well as a recall measure is obtained within the same test. The experimental design included a check on the possibility that differences in general intellectual ability might account for observed sex differences in social recall.

METHOD

Subjects

The *S*s were 103 males and 69 females in nine experimental groups at the University of Connecticut. The average group size was 19. Since the various tasks were administered at different sessions during the semester, the number of *S*s whose performances were available for analysis reflect some attrition because of absences. In general the *S*s appear to be representative of American college students as indicated in published norms on selected ability and personality tests.

GWU Memory for Names and Faces. On the George Washington University Memory for Names and Faces subtest (5) the *S* is asked to memorize the names under 12 picture-portraits of young men. Later the 12 pictures are presented again with 13 additional portraits. Two other untimed subtests intervene between the original presentation and the subsequent recall task. In the present study this interval was filled with the GWU Sense of Humor subtest and the temporal span was rigorously set at 20 minutes. From the matrix of 25 pictures *S* is required to associate the correct one of four pictures presented in multiple-choice form with each of the 12 names. Each correct multiple-choice association is scored 1 with a total maximum score of 12.

Miniature Life Situation. Five males and five females in each experimental group were participants in the miniature social situation. Each was assigned an alias by the group experimenter who conducted a very brief "sidewalk interview" for about one minute. In every case it was made certain that the participant *S* spoke his alias clearly at least once. Later *S*s in each experimental group were asked to match the correct name from the 10 aliases listed on the blackboard with the person who had adopted that name for the experiment. The participant *S*s were asked to stand in each group in a prescribed random order to be identified from one of the names on the blackboard. Although the 10 participant *S*s were different for each of the nine experimental groups, the 10 names used as aliases remained standard for all groups. A score of one was credited for each correct matching. The total maximum score was 10.

KW Memory for Names and Faces. An 11″ by 14″ photograph containing twenty 1¼″ by 1⅜″ picture-portraits of 10 college males and 10 college females randomly arranged by sex were given to each subject face down. Under each photograph was a printed name. Instructions were:

> Turn over the photograph. Study these pictures carefully. You will be asked to identify the names and faces later. You will have five minutes. The recall task will be very difficult so study the names and faces intently.

Exactly five minutes later each *S* was presented with a second 11″ by 14″ photograph containing thirty ⅞″ by ⅞″ picture-portraits: the original 20 plus 5 more student portraits of each sex. This set of 30 picture-portraits was also randomly arranged by sex, with numbers from 1 to 30 under each picture. The answer sheet accompanying this photograph consisted of 30 items. Each item contained five alternative name choices and a "foil" designation. All the name choices appeared on the first photograph. Instructions were:

> You studied 20 name-face combinations. The sheet you have contains 30 pictures. Obviously there are ten faces you have not studied. Five are males; five are females. These are the foils. Turn to *page two* of the exercise. First, circle the picture numbers of the foils. Pick exactly five males and five females. After you are satisfied with your choice for the foils, identify the remaining photographs. If you are unsure of an identification—guess. Leave no blanks.

Two scores for each *S* were derived from this test. A *total* score consisted of the sum of the number of

"foil" and name designations accurately identified. A *recognition* score was the sum of the "foil" choices marked correctly.

Test Administrations

The Modified Alpha Examination Form 9 and the GWU names-and-faces subtest were administered in different sessions early in the semester. About two months later, the miniature life situation task and the newly developed KW names-and-faces test were administered to the same *Ss* in one session for each of the nine experimental groups.

RESULTS AND DISCUSSION

For purposes of analysis the data from the nine different experimental sections were combined and treated homogeneously by sex groupings. Statistical analysis did not reveal significant differences between observers and participants in the miniature life situation. This finding was consistent with a similar analysis in a previous investigation (2). Three experimental sections yielded significantly higher scores than the other six on the miniature situation task, but this difference did not seem to be seriously contaminating because the sexes were equally represented in the aberrant sections. There were no significant differences between sections on scores obtained from the GWU and KW tests. The previous research had shown that pedagogical and instructional sets are not significant factors in the scores obtained on two social memory tasks.

The following results are based upon 103 males and 69 females in nine experimental sections:

1. Highly significant sex differences favoring females were obtained on all social memory tasks as demonstrated in Table 1.

2. Female superiority on the GWU and miniature situation social recall tasks was consistent with results from a prior investigation (2), although the magnitude of the differences, expressed in standard deviation units, was slightly smaller in this study. The percentage of females below the male mean was 30 for the GWU subtest, and 29 for the miniature situation task, almost identical with previous results.

3. The newly developed KW names-and-faces memory test is most sensitive to sex differences as demonstrated by the high *t* values in Table 1. The percentage of females below the male median was 21.7.

4. The mean age of the males was 19.88 with an *SD* of 2.34; for the females, 18.61 with an *SD* of 1.03. The *t* value of 4.26 suggested age as an influencing variable, but when age was correlated with total scores on the KW, the obtained product-moment coefficients were .102 and .139 for males and females, respectively. The null hypothesis is not rejected with these obtained correlation coefficients. The effect of the direction of these correlations, however, is to reduce rather than to increase sex differences.

5. The mean scores on the Modified Alpha Examination Form 9 were 108.29 for males and 108.19 for females. The correlation coefficients obtained between this intelligence test and the KW names-and-faces test were .01 and −.02 for males and females, respectively. Thus, intelligence test scores did not discriminate between the sexes, nor were they related to the newly developed KW instrument.

6. The suggestion from unpublished findings in the earlier study (2) that males remember males better than they remember females, and females remember females better than they remember males, was experimentally confirmed on the *KW* test in this investigation. Evidence for this tendency favoring like-sex recall is presented in Table 2. Standard scores with a mean of 50 and a standard deviation of 10 were employed in Table 2 for facility of comparison. Scores for the 69 males were combined with a random sample of 69 female scores. This population, thus equally distributed by sex, was used to derive transformation equations for male and female photographs on the KW, and for male and female participant-*Ss* in the miniature life situation. These transformations were then applied to the entire population. Standard score comparisons became necessary because *Ss* of both sexes found male pictures an easier task than female pictures in the KW, al-

TABLE 1

SUMMARY OF SEX DIFFERENCES ON THREE SOCIAL MEMORY TASKS

Tasks	Males ($N = 103$)		Females ($N = 69$)		t
	M	SD	M	SD	
KW Total	19.87	5.19	24.06	4.89	5.37***
KW Recognition	6.95	1.68	8.22	1.56	5.08***
GWU	9.31	2.42	10.01	1.92	2.11*
Life Situation	7.18	2.68	8.35	2.09	3.21***

* .05 level of confidence.

*** .001 level of confidence.

TABLE 2
ANALYSIS OF STANDARD SCORES BY SEX IN RECALL OF MALE AND FEMALE NAME-FACE COMBINATIONS

Task	Males (N = 103)				Females (N = 69)				t between Male & Female S's
	M	SD	t	r	M	SD	t	r	
KW Male Names-Faces	48.72	10.38			52.03	8.52			2.28*
			2.77**	.498			4.45***	.616	
KW Female Names-Faces	46.09	8.92			55.98	8.52			7.33***
Situation Male Names-Faces	49.15	10.14			51.19	9.72			1.32
			1.42	.517			1.93	.315	
Situation Female Names-Faces	47.72	10.75			53.54	7.30			4.23***

* .05 level of confidence.
** .01 level of confidence.
*** .001 level of confidence.

though male and female items were approximately of equal difficulty in the life situation. All relevant comparisons were made easier by this analysis and yielded the same significance levels as obtained by raw score analysis.

In the life situation there is some tendency to remember one's own sex, but only the female *S*s approach the .05 level of confidence. It should be noted, however, that the maximum raw score of the same-sex names-faces combinations in the life situation is only five. Hence, this approach is not a sensitive test of the trend considered here. The KW provided a better test of the tendency for same-sex recall because of the greater range in scores. The maximum raw score for same-sex names-faces combinations on this task is 15. In Table 2, probability levels of .01 and .001 for males and females, respectively, are apparent for the tendency to score better in recalling one's own sex rather than the opposite sex on the KW. It is interesting to note in Table 2 that females are superior to males in remembering male and female names-faces combinations on both KW and life situation tasks. This reaffirms the female superiority on total scores for three social memory tasks demonstrated in Table 1.

The large sex differences found on all three social memory tasks in this investigation were not reported in earlier studies (1, 4, 10) with similar tasks. The objective of constructing an instrument more sensitive to sex differences in social memory was successfully attained. The KW names-and-faces test devised by the present writers resulted in clearly different score distributions between the sexes. The distribution of male scores was bell shaped; the female score distribution was J shaped. While the difficulty level of the KW is appropriate for the males, it is much too easy for the females. An established effect is strongly apparent on the three social memory tasks, and factors related to this effect may be examined. Theoretical explanation of the results, however, must be *post hoc*.

The possibility that a general intellectual factor may account for the obtained sex differences is not likely. Previous investigations (1, 6, 8, 9, 10) have yielded very small to negligible relationships between the GWU names-and-faces subtest and intelligence or a memory factor in intelligence. This general finding is congruent with the present study in which the null hypothesis was not rejected with respect to intelligence, and with our earlier research (2) in which no relationship to digit span memory was obtained. Performance on social memory tasks thus seems to reflect attitudes, interests, and motivational variables: in short, personality characteristics rather than general intellectual functioning. It is likely that a perfect or near perfect score is within the capacity of most of our *S*s. Our tentative explanation for the consistent female superiority on the three social memory tasks in this experiment is found in the emphasis upon greater social facility for the female sex role with the concomitant development of better social skills.

The "same-sex recall" tendency in remembering names and faces was a provocative finding on both the miniature life situation and KW picture tasks. One simple explanation might be found in greater exposure to reference names and faces within each sex. The temptation to offer a "deeper" hypothesis must be resisted until more information about individual personality characteristics is available.

The present findings lend some indirect support to Reisman's (7) notion that there has been a continual shift in American character structure from "inner direction" to "other direction" during the past 50 years. In the present study 28 per cent of the males and 32 per cent of the females obtained perfect scores on the GWU names-and-faces subtest, while Moss reported (3) in 1926 that only one person in five hundred from a college population obtained a perfect score on the same subtest three decades earlier. The data are approximately comparable, and they reflect considerable improvement in social memory over a long period of time with two similar student populations.

SUMMARY

The results obtained from three different social memory tasks—one life situation and two names-and-faces picture tests—demonstrated clear superiority for 69 college females over 103 college males. Of the three tasks, a new names-and-faces picture memory test, designed as a more discriminating refinement of the George Washington University Social Intelligence names-and-faces subtest, yielded the largest sex differences. This new instrument, designated the KW Memory for Names and Faces Test, demonstrated a "same-sex recall" tendency within the population employed; namely, a tendency for males to remember male pictures better than they remember female pictures, and similarly for females to remember female pictures better.

Neither age nor sex differences in intellectual ability as measured by the Modified Alpha Examination Form 9 accounted for obtained sex differences in social recall. Correlations between these intelligence test scores and KW test scores for males and females were essentially zero.

REFERENCES

1. HUNT, T. The measurement of social intelligence. *J. appl. Psychol.*, 1928, **12**, 317–333.
2. KAESS, W. A., & WITRYOL, S. L. Memory for names and faces: A characteristic of social intelligence? *J. appl. Psychol.*, 1955, **39**, 457–462.
3. MOSS, F. A. Do you know how to get along with people? Why some people get ahead in the world while others do not. *Sci. Amer.*, 1926, **135**, 26–27.
4. MOSS, F. A., & HUNT, THELMA. Are you socially intelligent? An analysis of the scores of 7,000 persons on the George Washington University Social Intelligence Test. *Sci. Amer.*, 1927, **137**, 108–110.
5. MOSS, F. A., HUNT, THELMA, OMWAKE, KATHERINE T., & WOODWARD, L. G. *Social Intelligence Test.* (Rev. Form, 2nd ed.). Washington, D. C.: George Washington Univer., 1949.
6. RANDOLPH, JANE M. A statistical analysis of the Moss Social Intelligence Test. Unpublished master's thesis, Univer. of New Mexico, 1934.
7. RIESMAN, D. *The lonely crowd; a study of the changing American character.* New Haven: Yale Univer. Press, 1950.
8. THORNDIKE, R. L. Factor analysis of social and abstract intelligence. *J. educ. Psychol.*, 1936, **27**, 231–233.
9. THORNDIKE, R. L., & STEIN, S. An evaluation of the attempt to measure social intelligence. *Psychol. Bull.*, 1937, **34**, 275–285.
10. WOODROW, H. The common factors in fifty-two mental tests. *Psychometrika*, 1939, **4**, 99–108.

Received June 18, 1956.

METAPHORS FROM MUSIC CRITICISM[1]

ROGER W. BROWN, RAYMOND A. LEITER, AND DONALD C. HILDUM
Harvard University

IT is axiomatic with modern semantics that the associations between words and meanings are arbitrary rather than appropriate. To be convinced of the truth of this axiom we are advised to contrast equivalent terms from several languages. For some reason writers in English generally hit on the same trio: *horse*, *cheval*, and *Pferd*. Then follows the rhetorical question: "Is there anything about the animal in question that is more *horse*-like than *cheval*-like or *Pferd*-like?"

The only alternative to arbitrary association is usually thought to be phonetic symbolism, the sound that echoes its sense. A recent publication (2) offers evidence that some phonetic symbolism exists in natural languages. However, it is at best a minor linguistic phenomenon and the case for appropriate associations does not rest here. It is based on the nature of language growth. When someone invents a new machine or forms a concept or buys a dog or manufactures a soap powder his first thought is to name it. But these names are almost never arbitrary creations produced by juggling the sounds of the language into a novel sequence. We think hard and ask our friends to help us find a "good" name, a name that is appropriate in that its present meaning suggests the new meaning it is to have.

Sometimes new words are introduced by borrowing words or morphemes from classical languages. The biological sciences have been especially partial to this practice as *photosynthesis*, *streptoneura*, and *margaritifera* testify. In order to savor the appropriateness of these names a classical education is required and so, for most of us, they are functionally arbitrary. A much more common way to make a new name is by using words or morphemes already in the language; either extending the meaning of some word or recombining morphemes. A dog may be called *Spot* or *Rover*, a detergent may be called *Surf*, one kind of personality is said to be *rigid*. *Underwear*, *overcoat*, *railroad train*, *fireplace*, and *phonograph* are compounds that can easily be analyzed. Language grows by appropriate extension and this appropriateness sometimes enables us to anticipate the meaning of a new word or the extension of an old word.

An especially interesting class of appropriate names is that vocabulary of sensation which is also used to describe people and voice qualities and manners, such words as *warm*, *cold*, *bright*, and *dull*. Asch (1) wondered whether these words were also used to describe people in languages not historically related to English. He found that they were so used in such remote languages as Hausa, Malayalam, and Thai and is inclined to believe that there may be some common quality in thermal and personal warmth which makes the extended use appropriate and not an accident in the history of English. The rhetorical question now becomes: "Is there not something warm about the warm personality and something cold about the cold personality?"

The present research is concerned with the use of sense vocabulary to describe voice quality rather than persons. Music critics often write that a singer's tone is *dry* or *cold* or *white* or *brittle*. These words have familiar nonauditory meanings but their application to tone is sometimes said to be entirely incomprehensible. The critic may fancy that a voice can be *dry* or *white*, but the reader of his criticism has no idea what he means. In elementary composition courses students are sometimes warned against such flowery and uninformative locutions. We decided to work with *Ss* who were not familiar with the jargon of music criticism; to ask them to apply extended sensory words to recorded voices in order to learn how well they would agree on the denotative use of such terms. As Asch looked for agreement among unrelated languages, we decided to look for agreement among persons. If there should be such agreement we would take it as evidence that there was something "appropriate" in the extension of nonauditory vocabulary to voices. However, we were prepared to doubt that this "appropriateness" lay in some intersensory dimension common

[1] The authors are indebted to the Laboratory of Social Relations, Harvard University, for supporting this research.

to the narrow use of a term and to the voice quality it described. There might be another way to account for metaphorical extension.

METHOD

Selection of the Words

We began by examining the language actually used by music critics for the description of voice quality. We read the reviews of four critics until we had collected for each critic 40 words that were applied to voices and were not simply synonyms for *good* and *bad*. The critics and the dates of their reviews are:

1. Olin Downes, *The New York Times*, December 19, 1952 through April 30, 1954.
2. Irving Kolodin, *The Saturday Review*, April 10, 1954 through December 25, 1954.
3. Winthrop Sargeant, *The New Yorker*, May 22, 1954 through May 21, 1955.
4. George Bernard Shaw, *London Music*, May 28, 1890 through November 19, 1890.

It is amusing to note, in passing, that although Shaw wrote during the Golden Age of Opera when Melba, Patti, and the de Reszkes were singing in London he appears to have listened with a jaundiced ear. The terms most frequently used by the modern critics were nearly all complimentary, but not so with Shaw. He favored such terms as *screaming*, *yelling*, *brauding*, and *goat-bleating*.

We combined the four lists and ranked the words in order of total critical frequency. We then selected the 10 most common antonym pairs whose first dictionary meanings are sensory but not auditory. They are: brilliant-dull, closed-open, coarse-fine, cold-warm, dark-bright, hollow-full, light-heavy, smooth-rough, soft-hard, thick-thin. In addition, we selected the 20 most common words which do not have antonyms applicable to voice quality. Eleven of these words are primarily auditory in meaning and nine are not. The words are: brittle, chromium, dry, gravelly, hoarse, limpid, lustrous, lyric, nasal, pinched, reedy, resonant, ringing, round, shrill, solid, sonorous, strident, velvet, and white.

The Recorded Voices

We selected recordings of nine famous operatic voices; three Italian dramatic sopranos, three tenors, and three baritones. Within a vocal category the voices sang the same fraction of the same operatic aria. Each excerpt lasted approximately 30 seconds. The selections and singers follow:

Sopranos: *Vissi d'arte* from *Tosca* of Puccini. Played through the word *aiutai*.

1. Maria Meneghini Callas (S1).
2. Maria Caniglia (S2).
3. Renata Tebaldi (S3).

Tenors: *Che gelida manina* from *La Bohème* of Puccini. Played through the word *vicina*.

1. Mario del Monaco (T1).
2. Jan Peerce (T2).
3. Richard Tucker (T3).

Baritones: *Eri tu?* from *Un Ballo in Maschera* of Verdi. Played through the words *per me*.

1. Gino Bechi (B1).
2. Lawrence Tibbett (B2).
3. Leonard Warren (B3).

Subjects

The *S*s were 47 Harvard undergraduates who proved to be uninformed about opera and music criticism when asked a set of questions concerning these matters. They participated in groups of three, four, or five, but responded independently since all of their answers were written in test booklets.

Procedure

The *S*s began with a practice procedure involving a voice and words not used in the experiment proper. In the experiment, the order of presentation of the three voice categories (sopranos, tenors, and baritones) was randomly varied. Within a category, however, the voices were always presented in the same order for reasons of experimental convenience. To begin with, all three voices in a given category were played once. Recordings were played on an Admiral table model phonograph with an attached 12″ loudspeaker. Then the first voice was played again and *S*s marked the ten antonym pairs for that voice. For each pair there was a scale from one to four extending between the two adjectives, and *S*s circled the number best expressing their judgment of the voice in question. It might, for instance, be thought very brilliant—1, moderately brilliant—2, moderately dull—3, or very dull—4. Then the second voice was played and the pairs marked and finally the third voice. After the antonym pairs had been completed for the three voices, *S*s listened again to each one and picked the five words from the 20-word check list that seemed most aptly descriptive of the voice in question. The experiment continued in this fashion with the remaining two categories of voice. In summary, *S*s were required to judge each voice in terms of each antonym pair and to make free selection of the five most appropriate words from the 20 words listed without antonyms.

Finally, *S*s listened to the voices once again and indicated for each one whether or not they liked it. In five cases *S*s were privately interviewed after the standard procedure to learn how they would explain their word choices.

RESULTS

The Antonym Pairs

For our first test we have simply dichotomized the choices, combining ratings of 1 or 2 as choices of the adjective on the left, and ratings of 3 or 4 as choices of the adjective on the right. If *S*s had no basis at all for deciding which member of a pair to apply to a given voice, they might be expected to choose randomly. In that case the likeliest result for any pair with any voice would be an even division between the two words. A departure from chance expectation significant at the .05 level occurs with a division into 31 and 16 while for the .01 level the division must be 33 and 14, or more extreme. There are 38 differences significant at the .01 level and another 8 at the .05 level. In about half of the cases then, these

words were applied to voices with better than chance uniformity.[2]

It is possible that subjects favor one word in a pair over the other simply because the word is more familiar or because it is thought to be more suitable to the description of voices of whatever quality. Departure from chance does not demonstrate that the words actually suggest one kind of voice quality rather than another. However, critics use this vocabulary with the intention of suggesting particular qualities. We need to know whether the pattern of choices on a pair for one voice is significantly different from the pattern across the remaining eight voices. If choices on a given pair of adjectives divide in the same way on all nine voices then the words in question do not distinguish one voice from another. The test is chi square. The obtained frequencies of choices 1, 2, 3, and 4 for a particular pair on a given voice were compared with the frequencies to be expected on the basis of the distribution of choices for that pair across the remaining eight singers. Yates' correction was used in every case. A significant chi square indicates that the pair in question is applied to the singer in question in a manner different from the application of the same pair to the other singers. This is a test of the specificity of the semantic rule linking each voice with each adjective pair. It indicates the degree to which these adjectives distinguish one voice from the others. The significant chi squares appear in Table 1. Of the 90 chi squares computed (10 for each voice), 33 are significant at the .05 level or better, which is considerably more than the three or four to be expected by chance.

We also compared each of the three vocal ranges with the other two. Baritones, for example, were contrasted with tenors and sopranos to see if some adjectives showed a different choice pattern for baritones as a group than for tenors and sopranos combined into another group. The test again is chi square. The results make very good sense. Baritones are more likely to be called *dull*, *coarse*, *closed*, *dark*, *heavy*, *rough*, *hard*, and *thick*. This is at once a description of baritone quality and of the operatic characters played by baritones. Tenors are especially likely to be called *bright*, *light*, and *thin*, while sopranos are *coarse*, *soft*, *light*, and *thin*. There are more adjectives distinctive for baritones than for either tenors or sopranos and that, presumably, is because the baritone is a low register voice set against two high register voices. Tenors and sopranos are more alike on these dimensions than either is like the baritone. It was conceivable, of course, that the two male voices would resemble one another more closely than either would resemble the sopranos. It is fitting that this should not be so since tenor and soprano on the operatic stage invariably make common cause against the baritone.

[2] The complete table of choices for each antonym pair and for each voice has been deposited with the American Documentation Institute. Order Document No. 5136 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for 35 mm. microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

The Check List

From 20 words *S* chose 5 to characterize each voice. We wanted to know if a particular word was used more or less often for one voice than for the other eight. The test is again chi square with Yates' correction still necessary. We determined the number of times a particular word was applied to a particular singer and also the number of times it was applied to all other singers. Since there were 47 *S*s, a word always had the possibility of being applied 47

TABLE 1

THE SIGNIFICANT CHI SQUARES FOR EACH VOICE COMPARED WITH THE EIGHT OTHER VOICES ON THE ANTONYM PAIRS

Antonyms	Voices	χ^2 ($N = 47$)	Antonyms	Voices	χ^2 ($N = 47$)
Light-heavy	S1	12.16*	Coarse-fine	B1	26.63*
Thick-thin		29.51†	Cold-warm		40.25*
Coarse-fine	S2	12.67†	Closed-open		8.66*
Brilliant-dull	S3	44.63*	Dark-bright		8.11*
Coarse-fine		35.94†	Hollow-full		10.17*
Cold-warm		34.20†	Smooth-rough		34.62†
Closed-open		24.39†	Soft-hard		17.35†
Dark-bright		33.39†	Coarse-fine	B2	9.96†
Hollow-full		13.78†	Hollow-full		9.77†
Light-heavy		22.86*	Light-heavy		10.92†
Smooth-rough		31.37*	Thick-thin		11.07*
Soft-hard		38.74*	Closed-open	B3	7.89*
Hollow-full	T1	14.35*	Dark-bright		15.32*
Light-heavy		20.25*	Hollow-full		11.01†
Thick-thin		11.37†	Light-heavy		35.28†
Hollow-full	T2	14.33†	Thick-thin		76.05*
Thick-thin	T3	9.25†			

* The first word of the pair was more likely to be applied to the voice in question than to the other voices.

† The second word of the pair was more likely to be applied to the voice in question than to the other voices.

TABLE 2

THE SIGNIFICANT CHI SQUARES FOR EACH VOICE COMPARED WITH THE EIGHT OTHER VOICES ON THE WORDS FROM THE CHECK LIST

Word	Voice	χ^2	Word	Voice	χ^2
Shrill	S1	42.95*	Shrill	T2	13.59†
Brittle		9.80*	Brittle		5.58†
Round		8.15†	Pinched		6.84†
Solid		15.33†	Round		7.00*
Resonant		8.24†	Solid		8.33*
Shrill	S2	16.38*	Dry		9.61†
Pinched		3.89*	Resonant		4.87*
Sonorous		12.16†	Sonorous		4.06*
Gravelly		5.42*	Brittle	B1	30.04*
Strident		4.90*	Pinched		3.89*
Shrill	S3	4.41†	Round		4.32†
Brittle		6.00†	Dry		18.16*
Pinched		7.11†	Lustrous		8.57†
Dry		12.00†	Chromium		4.43†
Lustrous		10.81*	Lyric		16.42†
Reedy		4.17†	Nasal		11.04*
Sonorous		7.92*	Velvet		10.03†
Chromium		6.92*	Shrill	B2	12.57†
Gravelly		4.93†	Brittle		6.00†
Lyric		22.44*	Round		4.76*
Nasal		10.15†	Solid		18.32*
Strident		7.57†	Shrill	B3	8.01†
Velvet		26.90*	Solid		9.16*
Pinched	T1	8.00*	Reedy		5.94†
Lustrous		6.32†	Resonant		4.29*
Reedy		4.26*			

* The word in question was more likely to be applied to the voice in question than to the other voices.

† The word in question was less likely to be applied to the voice in question than to the other voices.

times to a particular singer and the possibility of being applied 376 (8×47) times to all other singers. The problem was to discover whether the attributions to one singer represented a larger or smaller proportion of the relevant possibilities than did the attributions to all singers. The significant results appear in Table 2. Of 180 chi squares 51 are significant with $p = .05$ or better. This is far more than the six or seven to be expected by chance.

Finally, we contrasted the frequency with which auditory words were chosen and the frequency with which nonauditory words were chosen. The mean number of choices of auditory words is 119 while it is 94 for nonauditory words. This difference does not attain to significance at the .05 level and so it is not clear that *Ss* found the auditory vocabulary more appropriate to the description of voices than the nonauditory vocabulary. Furthermore, among the auditory words there were no more cases of discriminating use than there were among the nonauditory. The two vocabularies do not differ in the frequencies with which they were used nor in the precision of that use.

DISCUSSION AND FURTHER ANALYSIS

We have asked *Ss* to apply sensory, nonauditory words to voice qualities. There was high agreement among the *Ss* on the pairing of some words with some voices. Furthermore, many of these words were used discriminatingly; i.e., the pattern of applications to one voice differed from the pattern for other voices. Suppose we knew that our *Ss* had experience of such words as *white*, and *dry*, and *thick* in their primary application to sense data but not in any extended auditory application. From the fact that the extended use shows agreement and discrimination we might conclude that the primary senses of *white*, and *dry*, and *thick* are concepts having appropriate application to voice quality. In effect, our *Ss* would have demonstrated that they possess shared, distinctive, supersensory concepts by their agreement on the identification of new auditory instances. The first difficulty with this conclusion is that we have not shown that our *Ss* have no prior knowledge of the extended auditory use of the words in question.

While it is likely that some of the terms were familiar designators of voice quality it seems likely that most of them were not. To begin with, of course, our *Ss* were not habitual readers of music criticism and were not knowledgeable about opera and opera singers. That fact alone does not guarantee that they were unfamiliar with the use of this vocabulary to describe voices. We must add that many *Ss* made complaints about the vocabulary offered them. They said that they would never freely choose such words as *white*, or *chromium* or *brittle* to describe voices. These complaints indicate that much of the vocabulary seemed unconventional to them and that our *Ss* had to rack their brains to find some reason for calling one voice *closed* and another *open*. The interview protocols confirm this impression. Especially instructive are the observations concerning the words *white* and *open*. In music criticism these terms name undesirable qualities. The *white* tone is colorless; the *open* voice is diffuse, lacking in focus. Many *Ss* used these terms as names of desirable qualities and interviews disclose how that happened. *White* can stand for purity, for a voice without blemish. *Open* can mean freely and easily produced. In these cases the words functioned as metaphors, as extensions of familiar terms to a new range of experience, extensions that were felt to be appropriate. The metaphor is most apparent

here because it runs counter to conventional critical use. The interviews indicate, however, that with most words *S* was unaware of any conventional usage and felt that he must discover reasons for selecting one word rather than another.

We believe it is reasonably clear that our *S*s felt themselves to be applying to voices a number of words that would not ordinarily be so applied. Furthermore, *S*s seem generally to have found one word more appropriate than another and to have agreed on which word it was. In short they appear to have created a number of metaphors and to have agreed on them. How do they come by the ability to perform in this way? Let us mention first the possibility with which this article began, a possibility suggested by Asch's study. It may be that when a child learns the sense meanings of words like *bright, dry,* and *rough,* he acquires a set of concepts applicable to voices and persons and manners. There is, however, another possibility that is suggested by factor analysis of our results.

The list of antonym pairs we used includes eight of the pairs used by Osgood and Suci (4) in their factor analysis of linguistic meaning. These experimenters had *S*s place 20 concepts (lady, boulder, sin, etc.) on 50 descriptive scales (good-bad, large-small, etc.). Analysis of the resultant matrix yielded three orthogonal factors which have been characterized as: (*a*) evaluation, (*b*) potency, (*c*) activity. The same factors have appeared in other studies and, from scales representing these factors, the Semantic Differential was constructed. Our *S*s placed voices (rather than concepts) on 10 scales defined by adjective antonyms. This procedure is clearly analogous to that of Osgood and Suci and yields data that can be factor analyzed. In addition, our *S*s applied 20 words (without opposites) to the same voices. These results were also included in the factor analysis. Finally, our *S*s were asked to say whether they liked or disliked each voice and those judgments have also been included.

The nine voices were ranked along each of the 31 scales according to the number of attributions they received from *S*s. It seems quite reasonable to assume that the numbers of attributions would be continuously and normally distributed over the universe of singing voices. However, because of the selected range of voices, we could not be sure of normality and it therefore seemed advisable to use only rank order correlations. In view of the very small *n* (9 voices) and the unorthodox use of rank-order *r* for factor analysis, the following results must be considered very tentative.

The first factor accounts for 56.7 per cent of the common factor variance. It may be called evaluation. The scales having factor loadings above .90 are *cold-warm, hollow-full,* and *like-dislike,* and the terms are *brittle, dry, lustrous,* and *pinched.* It is not possible to make detailed comparisons with the Osgood-Suci evaluation factor because our scales were not all the same. The scales most heavily loaded with their factor included *beautiful-ugly, sweet-sour, valuable-worthless,* and other pairs excluded from our study because of their obvious synonymity with *good-bad.* There is some indication, however, that the two evaluation factors would not be quite the same even if identical scales had been used. For instance, the scale *wet-dry* in the Osgood-Suci study is not loaded with the evaluation factor whereas the word *dry* is heavily loaded with evaluation in our study. This simply means that dryness is clearly undesirable in a voice but that, across 20 concepts, dryness is neither clearly bad nor clearly good. Changing the material to be judged somewhat alters the character of the evaluation factor but there is justification for calling both factors of evaluation, since a *good-bad* scale is central to each.

The second factor seems to be best characterized as potency. The third is more difficult to characterize, but there is some ground for thinking it an activity factor similar to that found by Osgood and Suci. We extracted a fourth factor, but find it impossible to characterize. It is interesting to note that these four factors account for over 80 per cent of the total variance while the Osgood-Suci factors accounted for only 50 per cent. With more restricted materials and scales we obtained more communality.

What is the origin of the semantic clusters revealed by the Osgood-Suci work and by the present study? Consider two scales from the potency factor; *large-small* and *heavy-light.* Imagine, as is quite possible, that these words were first encountered as names for attributes of a set of wooden blocks. If all the blocks were made of the same kind of wood the two scales would be in natural correlation; the larger the block the heavier it would be. The same sort of relationship binds *heavy-light* to *thick-thin* and to *wide-narrow.* Words of this kind are

ordinarily used to compare one object with others of the same type. A block is *large* relative to other blocks. As a consequence *large*, *heavy*, *thick*, and *wide* are terms that ordinarily imply one another. And the implications spread even more widely. When it is a question of people the larger, heavier, thicker, wider specimens seem to be more likely to have loud bass voices than are the smaller, lighter, thinner, narrower specimens. In short, sense qualities, though logically independent, are probably correlated in nature. Suppose, now, that *S* is asked to judge 20 concepts with respect to all these scales. A concept like "boulder" is referred to rocks and stones and, in comparison, judged to be *heavy*, *large*, *thick*, and *wide*. These terms are directly applicable to boulders. However, boulders have no voices. Where, then, does the concept belong on the *bass-treble* or *loud-soft* scales? We cannot doubt the answer. If Disney were to give a boulder a voice it would be *bass* and *loud* in contrast to the piping of a pebble. This could be a simple extrapolation of the usual correlation between these terms. It does not necessarily mean that there is anything in a bass voice *per se* that makes it appropriate to a boulder.

In our study *S*s may have felt that *thick* and *thin* simply do not apply to voices. However, *loud* and *resonant* do. Now thick people and animals and violin strings are likely to be *loud* and *resonant*, so if *S* is required to guess he will guess that *loud* and *resonant* voices ought to be called *thick* rather than *thin*. Similarly, *S* may feel that *cold* and *warm*, in their thermal sense, do not apply to voices. However, *warm* things are generally more agreeable than *cold* things and so a voice he likes is going to be called *warm*. Factor analysis suggests, then, that metaphors preserve correlations laid down in the earliest language experience. If *S* is asked to apply a word to a concept or experience to which it has no literal application he will make his decision so as to maintain the usual correlation between this word and others that do have literal application in the present instance. Karwoski, Odbert, and Osgood (3) have suggested that such natural correlations may also help to account for the clustering of sense qualities revealed by the study of synesthesia.

If there is a natural correlation in the use of words like *thick*, *heavy*, and *large* it does not follow, of course, that these words name the same dimensions. The Osgood-Suci findings must not be interpreted to mean that our lexicon is made up of countless perfect synonyms for *good-bad* and *strong-weak*. We have only to imagine a set of blocks that includes some made of balsa wood as well as some made of pine to realize that *heavy* and *large* name dimensions that are logically independent. However, Osgood and Suci lined up words with concepts to which they had no literal application. Forced to make some decision *S* applied the inapplicable words wherever he applied their usual correlates among the applicable words. So the words are not used with all the independence of which they are capable but, rather, with all the intercorrelation they generally manifest.

SUMMARY

A vocabulary of metaphorical terms used to describe voice qualities was drawn from the writings of four music critics. The *S*s who were not well informed about music listened to recordings of nine operatic voices and assigned terms from the metaphorical vocabulary to these voices. Many of the words were used with agreement and discrimination. Factor analysis of the results yielded three factors similar to those found in the more general factor analyses of linguistic meaning made by Osgood and Suci. The factors are: (*a*) evaluation, (*b*) potency, (*c*) activity. It is suggested that these factors may derive from correlations in nature between sense qualities. Metaphorical extension of the name of a sense quality may operate so as to preserve such correlations. Metaphorical extension of a sensory term need not be taken as evidence of an intersensory dimension.

REFERENCES

1. ASCH, S. On the use of metaphor in the description of persons. In H. Werner (Ed.), *On expressive language*. Worcester, Mass.: Clark Univer. Press, 1955.
2. BROWN, R. W., BLACK, A. H., & HOROWITZ, A. E. Phonetic symbolism in natural languages. *J. abnorm. soc. Psychol.*, 1955, **50**, 388–393.
3. KARWOSKI, T. F., ODBERT, H. S., & OSGOOD, C. E. Studies in synesthetic thinking: II. The role of form in visual responses to music. *J. gen. Psychol.*, 1942, **26**, 190–222.
4. OSGOOD, C. E. & SUCI, G. J. Factor analysis of meaning. *J. exp. Psychol.*, 1955, **50**, 325–338.

Received June 22, 1956.

LIKING FOR THE GROUP AND THE PERCEIVED MERIT OF THE GROUP'S BEHAVIOR[1]

LEONARD BERKOWITZ

University of Wisconsin

STUDIES consistently have shown that individuals tend to be more successfully influenced by groups to which they belong the more attractive group membership is to them (e.g., 1, 6). In many of these investigations attraction to the group has been defined in terms of the individual's liking for the other group members, but the findings are generalized to other sources of attraction as well (1). The major purpose of the present paper is to point to an effect of liking for the other group members that has been largely overlooked in theoretical interpretations of the relationship between liking and influence and to suggest that this effect may not occur for other sources of attraction to the group.

Perhaps the most fully developed theory of social influence is that proposed by Festinger (3, 5). Briefly, with regard to group attractiveness, this theory states that individuals are more highly motivated to achieve within-group consensus the more attractive the group is to them because pressures toward uniformity increase directly with this attractiveness (cf. Corollary VII A, 5, p. 131). These pressures supposedly are felt by the individuals themselves as well as being expressed in the behavior of the other members toward them (6).

Liking for the other group members—one source of attraction to the group—also can facilitate the group's influence over the individual members through an additional process. Liking can increase the perceived merit of the group's behavior in the absence of independent objective tests of this merit. Thus, subjects (*S*s) in these experiments may have conformed to opinions expressed by the other group members partly because, in liking these others, the *S*s perceived their opinions to be probably correct.

Something similar to the present hypothesis may be implied in several statements made by Festinger concerning the individual's acceptance of the group's opinions. He has proposed (3, 4) that there is a greater tendency for individuals to adopt the opinions advocated by the group the more attractive the group is to them. However, no explanation is given except to suggest that the attractive groups have greater "power" over their members (3), and there is no explicit statement that liking may affect the perceived merit of the other group members' opinions.

The major focus of this study, therefore, concerns the following hypothesis: *S*s with high liking for their other group members (in contrast to *S*s with lower liking for the others) (*a*) judge the opinions expressed by the others in the group as being more likely to be correct, and (*b*) show greater conformity to these opinions.

As a secondary purpose, the present experiment also investigates the effects of past proficiency upon influence. In line with findings of previous studies (9, 10) that *S*s are more readily influenced by others who previously have expressed many correct opinions than by those who have often been incorrect, it is reasonable to hypothesize that *S*s' evaluations of the merit of others' opinions are directly affected by these others' records of previous correctness on similar issues. One may then inquire whether liking and past proficiency combine to produce effects on the perceived merit of an opinion greater than that resulting from either factor alone, and to yield summative influence effects.

METHOD

Manipulation of Liking

Differences in *S*s' liking for their partners were obtained by employing a now standard procedure. At the time of volunteering to participate, students in introductory psychology discussion sections completed brief personality questionnaires on the basis of which they supposedly would be matched with a "congenial" partner. Two *S*s, both male or female, were scheduled for a given hour. When each *S* arrived to keep his appointment, he was ushered into a separate room for oral instructions. The *S*s in the *high liking* condition (High Like) were told that the matching had been very successful, that he and his partner should be very congenial and probably would like each other. Each of the

[1] The invaluable assistance of Mrs. Marsel Heisel, who collected the present data, is gratefully acknowledged.

Ss in the *low liking* condition (Low Like) was told that due to scheduling difficulties it was not possible to match him with a congenial partner, and that because the *S*'s personality and that of his partner were so dissimilar they might even grow to dislike each other.

Group Task

Following these instructions, the two *Ss* were brought together. *E* then explained that the purpose of the experiment was "to determine how people work together when they have to communicate by indirect means such as over a telephone system." The *Ss* were led to believe that they would represent an Army artillery team. One of the *Ss* would be the "observer" and the other the "battery commander." They were shown a series of charts supposedly representing the terrain lying before the "observer" at his forward post. The "observer's" task was to estimate the distance in inches from one point on the chart (his post) to each of a number of other points (the targets). It was explained that the "battery commander" would have two sources of information concerning the distances to the targets, the distance estimates telephoned to him by the "observer," and "an old and possibly unreliable map of the area" from which rough distance estimates were computed to each of the targets. The "battery commander's" task was to compare, for each of the large number of targets, the "observer's" estimate with the estimate from the "old map" and then write down the distance information to be used by his artillery.

Manipulation of Partner Proficiency

After the *Ss* returned to their individual rooms *E* told each *S* that he had been picked to be the "battery commander." The *S* was given a set of earphones by means of which, it was said, he could hear the "observer." He was also given a list of the 80 targets, the distances to each estimated from the "old map," and a sheet of paper on which he was to write his decisions. Each *S* actually worked alone and heard a previously recorded tape of the supposed interchange between *E* and the "observer." No *S* reported being aware that he was listening to a tape recording. The first part of the experiment consisted of a preliminary series of ten supposed practice trials for the "observer" which was employed in manipulating *Ss*' perceptions of the task proficiency of their "partners." The *Ss* were instructed that in the practice trials as throughout the entire task an estimate would be counted as a "miss" if it departed from the true distance by more than one inch. For each of the ten practice targets the supposed "observer's" voice reported a distance, after which the *E*'s voice was heard to say either "hit" or "miss." For the *high observer proficiency* condition (High Prof.), seven out of the ten "observer" responses were said to be "hits," while only three of the ten were termed "hits" in the *low observer proficiency* condition (Low Prof.).

While other evidence will be presented that the manipulation of the "partner's" perceived proficiency was at least partly successful, analyses of variance indicate that it was not completely so. The main effect for proficiency fell short of desirable standards of significance for *Ss*' judgments of the percentage of hits (correct opinions) by their "partners" in the main experimental series ($F = 3.18$, $p < .10$), and was nonsignificant for the behavioral conformity measure.

Procedure

When the preliminary series was completed, the main experimental procedure was followed for fifty distance estimates in five distinct series of ten targets each (Block 1). For each of these targets, *S* was to decide upon a distance to be written down using the two types of information supplied to him. The two distance estimates for each target ranged from being in complete agreement to four inches apart.

After the fiftieth target the *E*'s voice asked *S* to write down the percentage of targets on which he believed the "observer" had scored a hit. The procedure then was repeated for three more series of ten targets each (Block 2), and *Ss* again estimated the "partner's" percentage of hits. There were no reliable differences among the conditions in degree of change in the estimates from Block 1 to Block 2.

At the completion of the last series the tape recorder was stopped and the *Ss* were given a brief questionnaire. Three items, proven to be sensitive to the manipulation of liking in other experiments, were used as a scale to assess the liking of the *Ss* for their "partners." The items are as follows: 1. How well did you like your partner? 2. If you were taking part in another experiment, how much would you like to work with this same partner? 3. If it involved any effort on your part, would you care to know your partner better?

Conformity Measures

The distance decisions made by the *Ss* as "battery commanders" fell somewhere between the supposed "partner's" estimates and those derived from the "old map." Two conformity scores were arrived at by counting within each of the two blocks of series the total discrepancy (in inches) between the *S*'s decisions and the "partner's" estimates and then dividing each of these sums by the total possible discrepancy in the particular block, that between the values supplied by the "map" and the "observer." Both sets of percentages then were normalized by means of the arc sine transformation. The smaller the scores the more the *S*'s decisions had been influenced by his "partner."

Subjects

Analyses of the results with the scale measuring the *Ss*' liking for their partners indicated two sources of potential difficulty. First, the difference between the high and low liking conditions on this scale just fell short of significance (p approximately .10). Second, and perhaps more important, the *Ss* liked their partners more in the High Prof. condition than in the Low Prof. condition. As has been demonstrated in other studies (2, 8), individuals tend to be more attractive to other group members if they are seen as performing well than if they are seen as performing poorly. Since it was important for the purposes of the present experiment to eliminate this last difference, *Ss* were selected for inclusion in the final sample in such a manner that the liking scores were comparable in the High and Low Prof. conditions but divergent in the High and Low

Like conditions.[2] This selection resulted in statistically significant differences between the High and Low Like conditions ($p = .01$).

As a result of this screening, and where only the liking scores were known, twelve of the original 52 *Ss* were eliminated leaving ten *Ss* in each of the four conditions. Thirteen of the remaining *Ss* were females who were distributed approximately evenly among the conditions.

RESULTS

Perceived Correctness of the "Partner's" Opinions

Table 1 presents the condition means for the two estimations of the percentage of hits scored by the *Ss*' "partners." Since reliable individual differences in the judgments made by the various *Ss* were revealed by analysis of variance, the "*Ss* within conditions" term is employed as the error term.

The major hypothesis regarding the perceived merit of the "partner's" opinions is supported by the highly significant main effect for Liking ($F = 10.98$, $p < .01$). Liked "partners" are judged to have a higher percentage of hits (correct opinions) than are less liked "partners."

The hypothesis dealing with the combination of liking and past proficiency also is upheld. Because of the absence of significant differences between blocks of series, the t tests are based upon the condition means for all eight series. The *Ss* in the High Like, High Prof. condition report a significantly higher percentage of correct opinions for their "partners" than the *Ss* in the other conditions. Similarly, *Ss* in the Low Like, Low Prof. condition judge their "partners" as being correct significantly less often than the *Ss* in either the High Like, Low Prof. or Low Like, High Prof. conditions.

Conformity

Analyses of variance of the conformity scores based upon the combined eight series did not yield any significant F ratios, perhaps because of task withdrawal reactions in the Low Prof. condition. That the latter effect may have played a part is suggested by the fact that after the first series of targets, three of the *Ss* in the Low Like, Low Prof. condition had consistently written down as their decisions the exact estimates provided by either the "observer" or the "map," instead of compromising between these two sources of information as generally was the case.[3]

TABLE 1
MEAN PERCEIVED ACCURACY OF "PARTNER"

Time of Measurement	High Proficiency		Low Proficiency	
	High Like	Low Like	High Like	Low Like
After Series 1–5:	64.5	47.0	56.9	35.8
After Series 6–8:	62.3	45.0	49.8	42.0

TABLE 2
MEAN PER CENT OF POSSIBLE DISTANCE AWAY FROM "PARTNER" IN THE FIRST SERIES (ARC SINE TRANS.)

High Proficiency		Low Proficiency	
High Like	Low Like	High Like	Low Like
32.40	48.90	42.10	45.20

Because this type of extreme task reaction was not evident during the first series of targets, analyses of variance were conducted employing the normalized percentage data for this series. Nevertheless, the possibility remains that even at this time at least several *Ss* in the Low Like, Low Prof. condition were relatively indifferent to their assigned task of deciding between the two sources of information, so that even the first series data in this condition might be relatively unreliable. The condition means are given in Table 2.

The major hypothesis with regard to behavioral conformity is upheld by a significant main effect for Liking ($F = 11.11$, $p < .01$). Combining both proficiency conditions, *Ss* exhibit greater conformity to the opinions expressed by liked partners than to the less-liked partners' opinions.

The analysis of variance also revealed an unexpected interaction. The difference between the High and Low Like conditions is

[2] This selection is justified to a certain extent because we are interested in testing the effects of differences in liking, not whether the present instructions can successfully produce differences in liking. The present screening, of course, imposes a limitation upon the population to which the results can be generalized. A possibility remains that the selected subjects would uniformly be located toward the extreme end of a personality dimension that enters into the relationship between liking, perceived opinion merit, and conformity.

[3] It is worth noting that this presumed withdrawal from the task took place only in the Low Like, Low Prof. condition. The *Ss* in this condition did not have the satisfactions stemming from either liked or proficient "partners."

significant, by t test, for the High Prof. conditions ($p < .01$), but is nonsignificant for the Low Prof. conditions. Since *S*s in the Low Prof., High Like condition perceive their "partners" as having significantly more correct opinions than *S*s in the Low Prof., Low Like condition, it is possible that this nonsignificant difference in degree of conformity in the Low Prof. condition is an artifact of the hypothesized task withdrawal reactions.

The hypothesis dealing with the combined effects of liking for the "partner" and "partner" proficiency upon conformity to his opinions also is given some support. The *S*s in the High Like, High Prof. condition have a significantly greater tendency to conform to their "partners'" opinions than the *S*s in the other conditions. (One t value is significant beyond the .05 level, while the others are significant at less than the .01 level.) The other implication of this hypothesis, that *S*s in the Low Like, Low Prof. condition would have the lowest level of conformity, is not supported, again perhaps as a result of unreliability of the data in the Low Like, Low Prof. condition.

DISCUSSION

Recent theoretical interpretations of social influence explain the frequently observed high degree of conformity to the opinions advocated by attractive (liked) groups by postulating either that pressure toward uniformity are relatively strong in the attractive groups (3, 5) or, to take another recent theoretical model (7), that attractive groups have relatively great "power" over the individual members. Both of these formulations appear to imply some punishment for nonconformity. In contrast, the present hypothesis is that individuals also may conform to and adopt the opinions held by like groups because liking has increased the perceived merit of the group's opinions.[4] While the present experiment has not demonstrated unequivocally that the effect of liking for others upon the judged merit of the others' opinions produces the subsequent conformity to their opinions, the obtained findings generally appear to be consistent with this interpretation.

That the present hypothesis can account for some of the observed conformity behavior in the studies by Festinger and his colleagues also is suggested by the nature of their experimental situations. Typically, the *S*s in these experiments discussed issues about which they were probably not greatly concerned, without much time to think about the opinions they were asked to express. The *S*s in these experiments may have been uncertain about the correctness of their initial opinions, and, liking the others in their groups, have perceived the others' opinions as being better. Such uncertainty about one's initial opinion should be greater among opinion deviates, as Festinger, Gerard, *et al.* (6) have shown to be the case.

The present findings also have some implications for theoretical conceptions of the correlates of group attractiveness. Back has demonstrated (1) that attractive groups are more likely than unattractive ones to influence the behavior of the individual members, regardless of the source of attraction. It is not at all certain, however, that other sources of attraction than liking for the group's members also affect the perceived merit of the group's behavior or opinions.

REFERENCES

1. BACK, K. The exertion of influence through social communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 9–24.
2. BERKOWITZ, L., & LEVY, B. I. "Pride in group

[4] This hypothesis is similar to (but not identical with) that advanced by Newcomb (11) in his extension of Heider's general principle of "balanced states." According to Newcomb, *S*'s "strain toward symmetry" in a matrix consisting of himself, his orientation (attitude) toward B (*S*'s partner in the present sense), and his orientation toward X (an opinion advanced by B) increases with the valence of the objects of co-orientation. Therefore, the greater *S*'s attraction toward B, the more likely *S*'s orientation toward X will be "symmetrical" with his orientation toward B. However, Newcomb's hypotheses concerning the origin of this "strain toward symmetry" appear to prohibit the use of his model in the present situation. It is suggested that the "strain toward symmetry" arises as people learn the two advantages of symmetrical orientations: (a) ready predictability of another's behavior with regard to other Xs, and (b) validation of one's own orientation toward X. These advantages obviously do not hold in this case since the *S*s know they will not directly experience the Xs themselves, and there would be little reason for the present *S*s to achieve symmetry if these were the only benefits.

To tentatively advance another explanation for the present results, an individual who likes another associates this other with his self. Therefore, to the extent that he has a need to enhance the merit or value of his own self, he also will strive to enhance the merit of the liked person.

performance" and group task motivation. *J. abnorm. soc. Psychol.*, 1956, **53**, 300–306.
3. FESTINGER, L. Informal social communication. *Psychol. Rev.*, 1950, **57**, 271–282.
4. FESTINGER, L. An analysis of compliant behavior. In M. Sherif (Ed.), *Group relations at the crossroads*. New York: Harper, 1953.
5. FESTINGER, L. A theory of social comparison processes. *Hum. Relat.*, 1954, **7**, 117–140.
6. FESTINGER, L., GERARD, H., HYMOVITCH, B., KELLEY, H., & RAVEN, B. The influence process in the presence of extreme deviates. *Hum. Relat.*, 1952, **5**, 327–346.
7. FRENCH, J. R. P., JR. A formal theory of social power. *Psychol. Rev.*, 1956, **63**, 181–194.
8. GILCHRIST, J. C. The formation of social groups under conditions of success and failure. *J. abnorm. soc. Psychol.*, 1952, **47**, 174–187.
9. KIDD, J., & CAMPBELL, D. The conformity to groups as a function of group success. *J. abnorm. soc. Psychol.*, 1955, **51**, 390–393.
10. MAUSNER, B. The effect of one partner's success or failure in a relevant task on the interaction of observer pairs. *J. abnorm. soc. Psychol.*, 1954, **49**, 557–560.
11. NEWCOMB, T. M. An approach to the study of communicative acts. *Psychol. Rev.*, 1953, **60**, 393–404.

Received July 2, 1956.

STUDIES IN THE PHENOMENOLOGY OF EMOTIONS[1]

JACK BLOCK

Institute of Personality Assessment and Research, University of California

THE present paper describes an application of Osgood's "semantic differential" method (6), to the study of the phenomenology of emotions. Data dealing with sex and cultural differences are reported.

The semantic differential, as already attested by its widespread use in a variety of research enterprises, is a most versatile technique. By a standardized connotative description of various objects or concepts, the meaning of meaning is given what seems to be quite an appropriate quantification, making fruitful analyses possible.

The feature of interest here is that the method, when applied to the description of emotions, appears to provide operational access to the phenomenology of emotions. Phenomenologists, as for example Sartre in his small volume on emotions (10), have been persistent, and in the writer's opinion justified, in criticizing psychologists for neglecting the study of feelings. Although the criticisms have frequently stemmed from basically antiscientific philosophical positions, it does seem that psychology should be more concerned with the nature or meaning or conditions of private emotional experience. Some preliminary work suggested that the semantic differential could quantify affective qualities in an unstereotyped way and the studies reported here were therefore planned.

A word of caution first. There is always a degree of arbitrariness in positing an operation as a relevant or sufficient manifestation of a construct. What is required is that the arbitrariness be of a "reasonable" kind, where "reasonableness" unfortunately can be only vaguely evaluated. In suggesting that the procedure for description of emotions by the semantic differential could be taken as an operational index of the actual quality of an individual's emotional experience, two considerations were persuasive.

1. There is no dictionary definition of affects such as *humiliation*, *grief*, *boredom*, and so on, in connotative terms. Moreover, there is little, if any, institutional or cultural referent to help in the connotative description of emotions. We have not learned simple conventions to employ when the rather unusual task of describing emotions is presented to us. Indeed, as will be seen in Table 1 below, *envy* in actuality is described as somewhat more *red* than *green*, as the verbal cliche would dictate. How then is the task coped with by the individual? It would seem that in order to fulfill the demands of the test situation, the *S* can only turn inward for the basis of his report and what is inward is by implication his private emotional experience.

2. The more general properties of language habit systems, to the extent they do operate in the description of emotions, seem to reflect experiential qualities rather than initially fortuitous and phenomenologically irrelevant response tendencies. See, for example, the work on synesthesia (3, 9, 12) and physiognomic perception (12).

The proposed correspondence of construct and measure, of course, must receive its support from the network of results it engenders. The two studies reported here bear on the proffered interpretation, and it is to these we now turn. The specific expectancies underlying the present researches were:

a. As measured by the semantic differential, the phenomenology and interrelations of emotions in men are different from the phenomenology and interrelations of emotions in women.

b. Across cultures, differences in the phenomenology and interrelations of emotions are related to independently evaluated differences in cultural values and emphases.

STUDY I

The form of the semantic differential employed in the first study followed closely that described by Osgood (6). Fifteen emotions, selected *a priori* but with the intention of com-

[1] This investigation, part of a larger research project, was supported by research grant M-1078 from the National Institute of Mental Health of the United States Public Health Service. Mrs. Anne von der Lippe first recognized the interest of a cross-cultural comparison of phenomenologies. The translation of the description of emotions procedure, its administration in Norway, and many of the computations are her work. I am privileged to present her data here.

TABLE 1

RANK ORDERING OF CONNOTATIVE DIMENSIONS IN THE DESCRIPTION OF EMOTIONS (FEMALE SAMPLE)

Connotative Dimension	Emotion														
	Love	Anger	Envy	Humiliation	Nostalgia	Pride	Fear	Boredom	Contentment	Sympathy	Grief	Worry	Elation	Guilt	Anticipation
HIGH-low	4*	4	7	12.5	14	1	6	19.5	3	8	16	14	1	16	1
GREEN-red	11	20	13	14.5	3	13	15	4	8	5	5	10	11	12	13
WEAK-strong	20	18	14	7	6	20	10	3	13	11	5	7	17	4.5	19
ROUGH-smooth	12	3	6	10	18	10	4	12	15.5	20	9	3	12	4.5	10
ACTIVE-passive	5	1	4.5	11	20	2	5	18	10	13.5	14	5	2	15	2
EMPTY-full	19	13	9	4	7	19	11	1	17	18	7	9	19	9	20
SMALL-large	18	15	11	5	13	17	16	7	14	17	14	13	18	8	16
COLD-hot	15	16	10	14.5	8	15.5	8	5	12	16	3	8	14	7	15
CLEAR-hazy	9	6	12	17	19	3	14	15	6	-9	14	15	5	14	8
YOUNG-old	8	8	8	6	16	4	9	11	9	12	18	16	7	13	7
GOOD-bad	1.5	12	20	19	10	6	17	17	2	1	19	20	3	19	4.5
PEACEFUL-ferocious	6	19	18	16	2	11	20	9	1	2	12	17	8	17	11
SICK-healthy	17	10	4.5	3	12	15.5	7	6	18.5	15	5	4	16	3	17
ANGULAR-rounded	14	5	2	8	17	8	3	13	15.5	19	11	6	13	6	9
TENSE-relaxed	13	2	1	1	15	5	1	10	18.5	13.5	2	1	10	2	4.5
SAD-happy	16	7	3	2	4	14	2	2	20	7	1	2	20	1	18
SOFT-loud	7	17	16	9	1	18	18	8	7	3	8	12	15	11	14
WET-dry	10	11	15	12.5	11	12	12.5	14	11	10	10	11	9	10	12
BEAUTIFUL-ugly	3	14	19	20	5	9	19	16	4	4	20	19	6	20	6
FRESH-stale	1.5	9	17	18	9	7	12.5	19.5	5	6	17	18	4	18	3

* Low ranks mean that the capitalized end of the connotative dimension applies; high ranks, that the small-lettered pole is indicated.

prehensive coverage of the affective sphere, were used as the concepts to be defined. Each concept was evaluated in terms of twenty connotative 7-point scales. The emotions selected and the defining scales used may be read from Table 1.

The semantic differential form was administered to 40 male and 48 female college students in an undergraduate psychology course. The *S*s described the emotions by marking appropriate points along each defining scale. The *S*'s orientation was specifically to describe each affect as personally experienced, not as perceived in others or as cultural referents would suggest.

In order to compare the two groups, mean descriptions of each emotion were compiled for each sex separately. For a given emotion, the scores assigned to the twenty scales first were summed across the 40 males and, separately, across the 48 females. Then, for each emotion, the twenty sums for men, and the twenty sums for women were separately rank-ordered, to provide a quick comparability between descriptions.[2]

[2] *S*s, as seems introspectively appropriate, tended to emphasize middle intervals on certain affects such as *nostalgia*. Although the rank orderings of these affects are therefore of lower reliability because based on scores with low variability, no bias is therefore introduced into the correlations between affects.

In Table 1 are listed the rank orderings of the connotative scales for each of the fifteen affects, for the female sample. These data are presented here for their interest as a portrayal of the phenomenological qualities of the emotions considered. The data for the male sample are not presented because with the exception of the emotion of *grief*, the connotative descriptions of emotions by the men and by the women correlated very highly (see Table 4), the range being from .84 to .98, with a mean, computed via the *z* transformation, at .94. The rank-order correlation between the sexes for *grief* was only .66. What the meaning of this discrepancy may be is obscure, but for women *grief* is described as relatively *low*, *green*, *weak*, *smooth*, and *rounded*, while for men *grief* is relatively *high*, *red*, *strong*, *rough*, and *angular*.

The intercorrelations between the emotions were computed separately for both men and women. The matrix based upon the female sample is presented in Table 2. Because of the high correlations already noted between the male and female descriptions of emotions, the correlation matrix based upon the male sample is necessarily similar to that for the females. Enough latitude remains, however, so that although pair-wise, male and female descriptions of emotions are similar, the or-

TABLE 2
INTERCORRELATION OF EMOTIONS IN A FEMALE SAMPLE

Emotion	Emotion													
	Love	Anger	Envy	Humiliation	Nostalgia	Pride	Fear	Boredom	Contentment	Sympathy	Grief	Worry	Elation	Guilt	Anticipation
Love	—														
Anger	+.15	—													
Envy	−.48	+.68	—												
Humiliation	−.73	+.29	+.74	—											
Nostalgia	−.03	−.81	−.55	−.14	—										
Pride	+.68	+.62	+.17	−.36	−.63	—									
Fear	−.37	+.74	+.91	+.62	−.29	+.31	—								
Boredom	−.82	−.51	+.19	+.59	+.50	−.82	+.06	—							
Contentment	+.83	−.28	−.75	−.85	+.20	+.41	−.69	−.59	—						
Sympathy	+.66	−.40	−.70	−.57	+.59	+.15	−.63	−.24	+.73	—					
Grief	−.73	−.12	+.47	+.60	+.28	−.56	+.44	+.77	−.71	−.31	—				
Worry	−.67	+.38	+.81	+.73	−.22	−.24	+.80	+.47	−.87	−.62	+.81	—			
Elation	+.88	+.36	−.27	−.68	−.37	+.87	−.16	−.90	+.74	+.42	−.71	−.54	—		
Guilt	−.85	+.16	+.69	+.82	−.09	−.47	+.66	+.66	−.92	−.63	+.84	+.88	−.78	—	
Anticipation	+.82	+.56	−.04	−.48	−.48	+.92	+.07	−.91	+.55	+.26	−.65	−.36	+.92	−.61	—

ganization or interrelation of emotions within each sex may be quite different. For this reason, the two correlation *matrices* were correlated with each other.

A fundamental test of the functional equivalence of two sets of data is that there be an isomorphism of the relationships among the variables constituting each data set. That is, if there exists a strong relationship between emotions A and B in the one data set, there should also exist a strong relationship between these same variables in the second data set. The same pattern of covariation among variables must be present in the two corresponding matrices. An over-all method for characterizing the degree of isomorphism is simply to treat the correlations between variables as *scores* and to correlate all of these scores paired across data sets. The coefficient resulting from this inter-matrix correlation is thus an index of the similarity of covariation within the two sets of data.

The intercorrelation of the two matrices, computed over the 105 separate cells, was .93. A coefficient of this magnitude means that, as measured by the semantic differential, emotions interrelate in the two sexes with full equivalence. This finding is contradictory to initial expectation and suggests that to the extent the description of emotions is a measure of the quality of affective experience, the differences in expressed emotionality between men and women are not due to differences at the level of experience. Rather, these overt differences may be due to the learning of different behavioral patterns for the expression of emotions or to different sensitivities to particular presses.

A centroid method factor analysis was then carried out, starting from the matrix shown in Table 2, in order to dimensionalize the interrelationships among the emotions. Three factors appear to exhaust the meaningful variance. The original and rotated factor matrices are to be seen in Table 3. Factor rotation was accomplished by means of the analytic "quartimax method" (5), with a final 45° rotation of Factors I and II in order to increase the psychological meaning of the factors.

For two reasons, the results of the present factor analysis may be accepted as more reproducible than is usually the case. First, the fully equivalent correlation matrix based upon the male sample indicates that the identical unrotated factors would emerge in a second analysis. Second, the objective rotational method employed eliminates much of the ambiguity that otherwise complicates the interpretation of factor analytic results.

Factor I is defined at one end of the continuum by *contentment*, *love*, *elation*, *sympathy*, and *anticipation*, all of these being *joie de vivre* emotions. At the negative end of the dimension are found *guilt*, *worry*, *humiliation*, *grief*, *envy*, *fear*, and *boredom*, all of these being less happy conditions of man. The factor dimension involved here is clearly the *pleasantness-unpleasantness* polarity, the hedonic orientation

TABLE 3
ORIGINAL AND ROTATED FACTOR MATRICES
(All decimals omitted)

Emotion	Original Factor Matrix				Rotated Factor Matrix			
	I(a)	II(a)	III(a)	h^2	I(a)	II(a)	III(a)	h^2
Love	−888	344	203	948	862	396	221	949
Anger	215	912	065	882	−223	901	−143	882
Envy	772	583	144	957	−785	581	−059	957
Humiliation	872	060	062	768	−874	054	−042	769
Nostalgia	−136	−806	370	805	096	−712	538	806
Pride	−450	818	175	902	429	847	050	904
Fear	665	605	394	963	−703	657	192	963
Boredom	667	−688	113	931	−675	−665	181	931
Contentment	−967	−070	056	943	956	−036	170	944
Sympathy	−736	−350	457	873	684	−232	593	873
Grief	774	−336	390	864	−809	−267	368	861
Worry	894	188	255	900	−917	215	117	901
Elation	−776	587	102	957	760	612	060	956
Guilt	961	−071	108	940	−966	−070	020	939
Anticipation	−610	752	148	960	590	780	055	960

so fundamental for the individual's adaptation and behavior.

Factor II is defined by *anger*, *pride*, *anticipation*, *fear*, *elation*, and *envy* at one end of the continuum and by *nostalgia* and *boredom* at the other end. This dimension appears to reflect the extent to which the individual is organismically involved in the affect. It is not affective tone, as in Factor I, but affective intensity or *level of activation* (4) which is represented here.

Factor III reveals a most interesting and less anticipated ordering of the emotions. It is defined by *sympathy*, *nostalgia*, *grief*, and, to a lesser extent, *love*. The factor is unipolar and appears to encompass those emotions issuing from *interpersonal relatedness* and considerations of impermanence.

The first two of these factors are rather obvious and seem to correspond with those suggested by Schlosberg (11) in his analyses of emotion. The third factor cannot be linked to previous findings.

The present usage of the semantic differential differs from that of Osgood. Where Osgood correlates scales (connotative dimensions), the present approach correlates concepts (emotions). The difference in emphasis is analogous to that involved in the *R* versus *Q* distinction (2). The alternate approaches may lead to different results in the case where the pattern of covariation among scales is a function of the concepts employed. Osgood has already reported evidence for such concept-scale interaction (7).

It is perhaps for this last reason that the factor-analytic results of Study I can be related only in part to those reported by Osgood and Suci (8). The first two factors found in the present research can be linked roughly with the *evaluation* and *activity* factors of Osgood and Suci. But the *interpersonal relatedness* factor found here does not seem reconcilable with their *potency* factor or any of the residual factors they report.

STUDY II

The second study involved cross-cultural comparisons in the description of emotions. The description of emotions procedure was translated into Norwegian (with the exception of the *soft-loud* scale) and administered to 34 individuals in Norway.[2] Twenty of the group were male and fourteen were female. Because of the finding of no sex differences in the first study and because of the otherwise rather small *N*s available for analysis, these two groups were combined. There is, of course, a certain risk entailed by the cross-cultural extrapolation. Some age and socioeconomic differences between the Norwegian and American samples also exist in that the Nor-

[2] A copy of the Norwegian translation has been deposited with the American Documentation Institute. Order Document No. 5135 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for 35 mm. microfilm or $1.25 for 6 by 8 inch photocopies. Make check payable to Chief, Photoduplication Service, Library of Congress.

TABLE 4

CORRELATIONS BETWEEN DESCRIPTIONS OF EMOTIONS BY NORWEGIANS AND AMERICANS

Emotion	Norwegian-American Correlation	American male-American female Correlation
Love	.90	.95
Anger	.89	.97
Envy	.76	.95
Humiliation	.57	.84
Nostalgia	−.07	.86
Pride	.93	.88
Fear	.68	.95
Boredom	.81	.94
Contentment	.93	.98
Sympathy	.80	.92
Grief	.23	.66
Worry	.81	.95
Elation	.67	.95
Guilt	.88	.94
Anticipation	.52	.96

TABLE 5

THE TWENTY LARGEST CROSS-CULTURAL DIFFERENCES IN THE INTERRELATION OF EMOTIONS

Emotion	Inter-r among Norwegians	Inter-r among Americans
Grief-Guilt	.06	.83
Sympathy-Love	.88	.58
Worry-Grief	.20	.84
Nostalgia-Anger	.43	−.71
Elation-Love	.58	.88
Envy-Love	−.73	−.40
Pride-Love	.85	.57
Nostalgia-Love	.77	.08
Sympathy-Anger	.31	−.47
Elation-Anticipation	.35	.91
Anticipation-Pride	.31	.88
Pride-Envy	−.44	.36
Nostalgia-Pride	.55	−.61
Fear-Anger	.19	.79
Sympathy-Pride	.67	−.11
Sympathy-Guilt	−.86	−.59
Sympathy-Contentment	.96	.74
Boredom-Elation	−.50	−.91
Grief-Envy	−.29	.59
Grief-Contentment	−.03	−.79

wegian sample is slightly older over-all, and with greater age range, and averages somewhat lower in social class. To the extent these several differences are relevant, the comparisons across cultures should show lack of correspondence, and interpretation of obtained differences becomes more difficult.

The data from the Norwegian sample were treated as were the American data. For each emotion, the correlation between the description by Norwegians and the description by American males was then computed. These may be read from Table 4. With the exception of *grief* and *nostalgia*, the correlations are reasonably high.

The intercorrelation matrix of the 15 emotions was derived, and this matrix in turn correlated with the matrix for the American males, in order to evaluate the extent of equivalence in the way emotions are *organized* in the two cultures. The inter-matrix correlation coefficient was .69 (.70 when three emotions were deleted for which there were *a priori* reservations as to the adequacy of the translation). A correlation of this magnitude is indicative of an appreciable degree of congruence between the two groups, particularly in the light of the potentially attenuating factors previously mentioned. As measured by the semantic differential, it would appear that emotions stand in roughly the same relation to each other in the American and Norwegian cultures. What the interrelations may be in other cultures remains to be assessed, but it does not seem an unlikely hypothesis that the universal experiences of man generate a common phenomenological background. Some additional evidence for this interpretation is to be found in a recent study by Asch (1).

Although the two cultures correspond greatly in the way their emotions are organized, the original interest was in the nature and meaningfulness of the differences observed. To pursue this concern, the twenty greatest differences in the way emotions were related in the two cultures were selected. That is, for a given pair of emotions, e.g. *grief-guilt*, the correlation in one culture was .83 and in the other culture .06. Differences approaching this magnitude account for the lack of total congruence between the two matrices.

The twenty pairs of emotions chosen are listed in Table 5. If the differences between cultures are meaningful, it was hypothesized, judges with some knowledge of both cultures should be able to predict the nature and direction of the differences found. Sixteen judges, psychologists and psychiatrists, attempted the prediction task, with the alternatives randomly arranged. Where chance would offer only 10 correct choices, the judges' consensus was accurate in 17 of the 20 estimates, an occurrence significant beyond the .01 level. For the judges individually, ac-

curacy scores ranged from 8 to 16, with a mean of 13.8. The differences between the two groups, despite the strong congruence in the way they depict emotions, do appear to be meaningful when evaluated in this fashion.

What these differences suggest, so far as cultural emphases and values are concerned, is open to varying interpretation. The adequacy of a crosscultural translation can perhaps always be questioned. There is the possibility, slight in the writer's estimation, that the differences observed reflect class distinctions rather than cultural differences. And certainly a number of more psychological hypotheses present themselves. For example, the strong relationship between *grief* and *guilt* in the American culture where they are orthogonal among Norwegians has psychoanalytic import and suggests a confounding of the two emotions by Americans. But it is probably wisest at the present stage of information to delay interpretation until further research evidence is in.

SUMMARY

Two studies are reported employing the semantic differential as a means of quantifying the phenomenology of emotions. In the first study, American men and women college students were found to describe emotions connotatively in a highly similar way. Factor analysis of these data revealed three underlying dimensions: a *pleasantness-unpleasantness* factor, a *level of activation* factor, and a factor derived from emotions involving *interpersonal relatedness*.

The second study compared a sample of Norwegians with Americans with respect to their descriptions of emotions. Although an appreciable congruence was observed, the differences in phenomenological organization proved to be predictable and hence presumably reflective of differences in cultural values.

REFERENCES

1. ASCH, S. On the use of metaphor in the description of persons. In H. Werner (Ed.), *On expressive language*. Worcester, Mass.: Clark Univer. Press, 1955. Pp. 29–38.
2. BLOCK, J. The difference between *Q* and *R*. *Psychol. Rev.*, 1955, **62**, 356–358.
3. KARWOSKI, T. F., ODBERT, H. S., & OSGOOD, C. E. Studies in synesthetic thinking: II. The role of form in visual responses to music. *J. gen. Psychol.*, 1942, **26**, 199–222.
4. LINDSLEY, D. B. Emotion. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951, Pp. 473–516.
5. NEUHAUS, J. O., & WRIGLEY, C. The quartimax method: an analytic approach to orthogonal simple structure. *Brit. J. statist. Psychol.*, 1954, **7**, 81–92.
6. OSGOOD, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1952, **49**, 197–237.
7. OSGOOD, C. E. Report on development and application of the semantic differential. Unpublished manuscript, Inst. of Communications Research & Dept. of Psychol., Univer. of Illinois.
8. OSGOOD, C. E., & SUCI, G. J. Factor analysis of meaning. *J. exp. Psychol.*, 1955, **50**, 325–338.
9. RIGGS, L. A., & KARWOSKI, T. Synaesthesia. *Brit. J. Psychol.*, 1934, **25**, 29–41.
10. SARTRE, J.-P. *The emotions: outline of a theory*. New York: Philosophical Library, 1948.
11. SCHLOSBERG, H. Three dimensions of emotions. *Psychol. Rev.*, 1954, **61**, 81–88.
12. WERNER, H. *Comparative psychology of mental development* (Rev. ed.), Chicago: Follet, 1948.

Received July 3, 1956.

EFFECTS OF BIOLOGICAL AND SOCIAL MOTIVATION ON SCHIZOPHRENIC REACTION TIME

GERALD ROSENBAUM, WILLIAM R. MACKAVEY, AND JAMES L. GRISELL

Wayne State University and Northville State Hospital

THEORY and research on the nature of schizophrenia appear to have been converging in recent years toward an attempt to derive disturbances in schizophrenic performance from a central motivational pathology. Hoskins (9), Babcock (3), and others (1, 15) have accumulated evidence pointing to a deficit in the basic biological motivation and the innate equipment necessary for coping with environmental stresses. On the other hand, Sullivan (19), Cameron (6), and Hunt and Cofer (10), among others (14, 17, 21), have marshaled support for the intactness of the schizophrenic's capacity for biological arousal. They have suggested that characteristic disturbances in schizophrenic language, thought, attention, and affect may be most accurately regarded as extensions of an acquired deficit in the social drives, intensified during the process of social adaptation.

A recent study by Cohen (7) was designed to investigate this issue by comparing schizophrenic performance under biological motivation with performance under ordinary social incentives. Finding that shock induced motivation improved motor learning more in schizophrenics than in normals, Cohen suggested that primary motivation and reward may be employed to compensate for inadequate social motivation in effecting learning in schizophrenics. The present experiment was designed to test the related hypothesis that characteristic disturbances in schizophrenic performance can be eliminated under biological shock motivation.

A task which has consistently yielded decrements in schizophrenic performance is the reaction time procedure originally employed by Huston, Shakow, and Riggs (12) and later developed into a schizophrenic set index by Rodnick and Shakow (18). Several repetitions of the original study under slightly different conditions (8, 11, 13, 20) have confirmed the essential findings that schizophrenics are slower, more variable, and have more difficulty in maintaining a regular preparatory set than normals. If such disturbances in schizophrenic attention and performance can be accounted for by impaired social motivation, then the arousal of a biological drive such as pain should improve schizophrenic performance to a normal level.

METHOD

To test this hypothesis, schizophrenic and normal *Ss* were tested on a reaction time task, in which a buzzer signal was presented at several regular and irregular preparatory intervals. Half the *Ss* in each group were then immediately retested on the same task under biological motivation, activated by a shock emitted by the response key at the same instant that the buzzer signal occurred. The other half of each group was retested as a control for the effects of practice without shock. It was predicted that schizophrenics would show significantly more improvement than normals under shock conditions, thereby reaching the same level as the normals in reaction time, variability, and ability to maintain a regular set.

Subjects

A group of 32 schizophrenic patients and a group of 32 employees from Northville State Hospital, matched for age and sex, were each divided into an experimental and a control group. Each of these four groups contained 10 males and 6 females. The mean ages in years for the four groups were as follows: schizophrenic-experimental 33.9, range 17–52; schizophrenic-control 32.3, range 24–47; normal experimental 34.9, range 24–52; and normal-control 33.9, range 24–52. All patients had been hospitalized less than three years.

Apparatus

The stimulus presentation and recording units of the reaction time apparatus were housed in a portable wooden cabinet 20.5 in. long, 17 in. wide, and 10.5 in. high. The *S*'s response unit consisted of a telegraph key and copper plate mounted on the far end of a 22 x 17 in. flat panel, hinged to the cabinet. The illumination of a 25-w. bulb, centered on the front face of the cabinet 8 in. above the panel, served as the ready signal.

The ready signal and a Hunter Decade Interval Timer were simultaneously activated by *E*'s closing a push-type switch. Depression of the telegraph key by *S* in response to the ready signal closed a holding relay, permitting *E* to release his switch. At the termination of the interval previously set on the interval timer, a door bell buzzer and Standard Electric Timer were instantaneously activated. Release of the response key to the buzzer signal broke the common circuit to the

TABLE 1
REACTION TIME SCORES IN .01 SEC. OF SCHIZOPHRENICS AND NORMALS

	Schizophrenics				Normals			
	Experimental		Control		Experimental		Control	
	M (N = 10)	F N = (6)	M (N = 10)	F (N = 6)	M (N = 10)	F (N = 6)	M (N = 10)	F (N = 6)
Initial Test								
Mean	35	39	49	79	23	25	24	24
Median	32	37	31	53	23	25	24	22
Retest								
Mean	19	33	37	93	17	19	22	24
Median	17	29	31	40	17	20	22	22

TABLE 2
VARIABILITY SCORES OF SCHIZOPHRENICS AND NORMALS

	Schizophrenics				Normals			
	Experimental		Control		Experimental		Control	
	M (N = 10)	F (N = 6)	M (N = 10)	F (N = 6)	M (N = 10)	F (N = 6)	M (N = 10)	F (N = 10)
Initial Test								
Mean	158	162	243	584	69	95	66	70
Median	97	153	119	346	61	94	66	64
Retest								
Mean	64	137	109	467	46	60	49	54
Median	43	83	91	211	44	54	42	55

buzzer and clock. Reaction times were read directly to the nearest .01 sec.

The shocks, generated by a transformed 50-v. current in parallel with the buzzer and clock, were administered through a copper electrode on the surface of the response key and the copper plate on which S rested his hand. The setting of a toggle switch determined whether or not shock would accompany the buzzer.

Procedure

All Ss were initially tested under the social instructions that they were taking a reaction time test requiring them to depress the key to the light signal and to release it as rapidly as possible at the sound of the buzzer. Each S was then given 10 practice trials at each of the three preparatory intervals, 2, 5, and 10 sec., first presented in that order in regular homogeneous blocks and then in an irregular randomized order. The first two reactions at each of the three regular intervals and the first six reactions of the irregular procedure were counted as practice trials and excluded from all computations, leaving 48 recorded test trials.

Following this initial run of 60 trials, a 5-min. rest was introduced to reduce fatigue effects. The control group of schizophrenics and the comparable normals were then retested under the identical conditions of ordinary social motivation. Both experimental groups were also retested on the same procedure but were instructed that they would "... feel some skin stimulation on the key at the same moment the buzzer sounds." Shocks were administered to the experimental groups on all 60 retest trials. All anticipatory responses were immediately repeated, so that every S was required to react to the buzzer on 120 trials.

RESULTS

The major results of the experiment are presented in Tables 1 and 2, which show the reaction time and variability scores of the schizophrenic and normal groups, broken down for sex, on both the initial test and retest. The individual reaction time scores on which the group data in Table 1 are based were obtained for each S by computing the overall mean of his 48 response latencies to the three test intervals in the regular and irregular orders. The variability scores comprising the basic data for Table 2 were computed for each S by obtaining the range of his eight reaction time scores for each test interval in the regular and irregular procedures, and then summing these six ranges. In view of the marked variability of the schizophrenic groups, medians as well as means are shown in the tables to permit an evaluation of the group trends relatively independently of the extreme scores.

Social Motivation

Support for the hypothesis that the schizophrenics showed an initial deficit in social motivation is afforded by examination of the first two rows of Tables 1 and 2, which show that all four schizophrenic groups were consistently slower and more variable than the comparable normal groups. To evaluate these trends statistically, all scores were transformed to the reciprocals of their square roots so as to eliminate the heterogeneity of variance occasioned largely by the extreme variability of the schizophrenic groups. Analysis of variance of the transformed reaction time scores for the initial test showed that the schizophrenics were significantly slower than the normals (.01 level), and that females were significantly slower than males (.05 level). No significant differences were obtained between experimental and control groups, nor were any of the interactions between the three main sources of variance significant. This finding indicates that there was no appreciable selection bias in assigning *Ss* to experimental and control groups. Similar analysis of the initial test variability scores showed that the diagnostic category was the only significant source of variance (.01 level).

TABLE 3

SUMMARY OF ANALYSIS OF VARIANCE OF TRANSFORMED VARIABILITY SCORES ON RETEST

Source	*df*	Mean Square	*F*
Diagnosis (D)	1	.0713	9.90**
Motivational Condition (M)	1	.0539	7.49**
Sex (S)	1	.0439	6.09**
D × M	1	.0319	4.43*
D × S	1	.0062	—
M × S	1	.0042	—
D × M × S	1	.0001	—
Within Groups	56	.0072	
Total	63		

* Significant beyond the .05 level.
** Significant beyond the .01 level.

TABLE 4

MEAN REACTION TIMES IN .01 SEC. AT DIFFERENT PREPARATORY INTERVALS

Experimental Group	Initial Test			Retest		
	Interval in Sec.			Interval in Sec.		
	2	5	10	2	5	10
Shizophrenic Exp.						
Regular	38	34	32	21	26	23
Irregular	45	38	32	28	24	24
Schizophrenic Control						
Regular	57	64	70	61	51	58
Irregular	67	51	52	60	63	59
Normal Exp.						
Regular	24	23	24	16	17	18
Irregular	28	23	23	21	19	18
Normal Control						
Regular	25	23	23	21	21	22
Irregular	27	23	22	26	23	23

Biological Motivation

The differential effects of biological and social motivation on the schizophrenic and normal groups are shown in the retest data of Tables 1 and 2. Inspection of these group trends reveals that the experimental groups receiving shock tended to be faster and less variable than the control groups, and that this tendency was more marked in the schizophrenic groups. Transformation and subsequent analyses of variance of these scores showed that the normals, experimental groups, and males were all faster and less variable than their comparable groups at the .01 level of significance. The results of the analysis of variance of the variability scores are presented in Table 3.

Since both schizophrenics and normals showed improvement in performance with shock, the interaction between these two main effects constitutes a test of the hypothesis that the experimental groups of schizophrenics improved over their comparable controls more than did the normals. Table 3, and careful scrutiny of the retest data in Table 2, reveal that shock did significantly reduce variability in both male and female schizophrenics, but produced only a slight decrease in the variability of normals. While this same trend seems to appear in the reaction time data in Table 1, it appears less marked, and statistical analysis revealed no significant interaction. None of the remaining interactions with the sex factor proved to be significant for either the reaction time or variability data.

Preparatory Set Data

The principal findings with regard to the effects of regular and irregular preparatory sets are presented in Table 4. While other investigators (11, 12, 13, 18, 20) have consistently found that normals react more rapidly with the regular than with the irregular procedure,

no reliable difference between these two procedures is apparent in either of the normal groups. Since the normal *Ss* were not appreciably faster on the regular than the irregular procedure, the effect of biological and social motivation on the typical schizophrenic failure to benefit from the regular preparatory set could not be investigated.

Further inspection of Table 4 indicates that only one preparatory set phenomenon was observed consistent with previous findings: the longer latencies on the 2-sec. irregular interval. These slow reactions appear to result from the inability of *Ss* to anticipate the trials requiring a brief response set. The present failure to replicate previous observations on preparatory set appears to be an artifact of procedural modifications, introduced because the present retest addendum made it desirable to shorten and to simplify the testing procedure. These modifications included fewer test trials at each interval, administration of the regular procedure first rather than last, and presentation of the regular intervals in the order 2, 5, 10 rather than 5, 2, 10. These modifications, incidentally, should tend to reduce the facilitating effects of warm up and practice on the temporal conditioning phenomenon inherent in the regular procedure.

DISCUSSION

The present results confirm previous findings that schizophrenics are slower and more variable than normals when performing a reaction time task under ordinary social motivation. Confirmation of previously reported sex differences favoring males (8) and a finding that females have higher variability scores than males were also obtained.

The major hypothesis, that schizophrenics would improve more than normals with shock motivation, was clearly supported for the variability scores but not for average reaction time. Evidence that schizophrenics can achieve a normal level of functioning when biologically motivated, however, can be seen in the data for males in Tables 1 and 2. The median variability and reaction time scores of schizophrenic males in the experimental group are clearly similar to those of normal males under shock motivation. These results are consistent with Cohen's findings for motor learning (7) and those of Paintal (16) for the GSR, supporting the view that the biological motivational systems of schizophrenics are unimpaired and when activated may greatly reduce the characteristic deficit in schizophrenic performance.

The very marked individual differences exhibited by schizophrenic *Ss* suggest, however, that this generalization does not hold for all schizophrenics. While the group trends were in the appropriate direction, two of the schizophrenic males and three of the schizophrenic females displayed mean reaction times under shock motivation that were slower than those of all comparable normals under ordinary social motivation. Furthermore, one of these schizophrenic females showed an *increase* in mean reaction time from .36 seconds without shock to .58 seconds with shock.

Clinical evaluation of these atypical individuals in the present schizophrenic sample indicated chronicity of their symptom patterns. This observation suggests that defective biological motivation may be characteristic either of the more chronic forms of schizophrenia or of some late stage of the schizophrenic process (2). This suggestion is consonant with Bellak's proposal (4) that schizophrenia may be a reaction type associated with varying biological and sociopsychological predispositions, as well as with Brackbill's recent suggestion (5) that some kind of structural brain damage may be found in those schizophrenic patients who show the classical signs of formal thinking disorder. Investigation of the relationship of the presently obtained index of defective biological motivation to variables such as age, length of hospitalization, duration of illness, presence of thinking disorder, and extent of deterioration, should facilitate an improved conceptualization of schizophrenia and the relevant dimensions of motivation.

SUMMARY

The present experiment was designed to test the hypothesis that characteristic disturbances in schizophrenic reaction time are the result of defective social motivation, and can be eliminated under biological shock motivation.

Thirty-two schizophrenic patients and 32 hospital employees, matched for age and sex, were first tested on a reaction time task in which they were instructed to lift their finger from a telegraph key in response to a buzzer presented 10 times at each of three regular and three irregular preparatory intervals. Half the

patients and normals were then immediately retested on the same task under biological shock motivation, and the other half of each group was retested as a control for the effects of practice without shock.

The results showed that the normals and the males were faster and less individually variable than the schizophrenics and the females under all conditions. Shock was also significant as a main source of variance, resulting in greater improvements in speed and variability than did practice. No difference between the regular and irregular orders was obtained for any of the groups or subgroups.

The major hypothesis was partially supported in that schizophrenics showed significantly greater improvement with shock than did normals, and shock motivation raised the median scores of the males to the level of the normals. However, the female schizophrenics and some individual males did not improve significantly with shock.

It was concluded that disturbances in schizophrenic reaction time are only partially accounted for by defective social motivation and it is suggested that a biological deficit may be involved in the reactions of more chronic schizophrenics.

REFERENCES

1. ANGYAL, A., FREEMAN, H., & HOSKINS, R. G. Physiological aspects of schizophrenic withdrawal. *Arch. neurol. Psychiat.*, 1940, **44**, 621–626.
2. ARIETI, S. *Interpretation of schizophrenia*. New York: Robert Brunner, 1955.
3. BABCOCK, HARRIET. *Dementia praecox: a psychological study*. Lancaster, Pa.: Science Press, 1933.
4. BELLAK, L. *Dementia praecox*. New York: Grune & Stratton, 1948.
5. BRACKBILL, G. A. Brain dysfunction in schizophrenia. *Psychol. Bull.*, 1956, **53**, 210–226.
6. CAMERON, N. The functional psychoses. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944, Pp. 861–921.
7. COHEN, B. D. Motivation and performance in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, **52**, 186–190.
8. HALL, K. R. L., & STRIDE, E. Some factors affecting reaction times to auditory stimuli in mental patients. *J. ment. Sci.*, 1954, **100**, 462–477.
9. HOSKINS, R. G. *The biology of schizophrenia*. New York: Norton, 1946.
10. HUNT, J. McV., & COFER, C. N. Psychological deficit. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944, Pp. 971–1032.
11. HUSTON, P. E., & SENF, RITA. Psychopathology of schizophrenia and depression: I. Effect of amytal and amphetamine sulfate on level and maintenance of attention. *Amer. J. Psychiat.*, 1952, **109**, 131–138.
12. HUSTON, P. E., SHAKOW, D., & RIGGS, L. A. Studies of motor function in schizophrenia: Reaction time. *J. gen. Psychol.*, 1937, **16**, 39–82.
13. HUSTON, P. E., & SINGER, MARY M. Effect of sodium amytal and amphetamine sulfate on mental set in schizophrenia. *Arch. neurol. Psychiat.*, 1945, **53**, 365–369.
14. MALMO, R. B., SHAGASS, C., & SMITH, A. A. Responsiveness in chronic schizophrenia. *J. Pers.*, 1951, **19**, 359–375.
15. MAY, P. R. A. Pupillary abnormalities in schizophrenia and during muscular effort. *J. ment. Sci.*, 1948, **94**, 89–98.
16. PAINTAL, A. S. A comparison of the galvanic skin responses of normals and psychotics. *J. exp. Psychol.*, 1951, **41**, 425–428.
17. RICKERS-OVSIANKINA, MARIA. Studies on the personality structure of schizophrenic individuals: Reactions to interrupted tasks. *J. gen. Psychol.*, 1937, **16**, 179–196.
18. RODNICK, E. H., & SHAKOW, D. Set in the schizophrenic as measured by a composite reaction time index. *Amer. J. Psychiat.*, 1940, **97**, 214–225.
19. SULLIVAN, H. S. *Conceptions of modern psychiatry*. Washington, D. C.: William A. White Foundation, 1947.
20. TIZARD, J., & VENABLES, P. H. Reaction time responses by schizophrenics, mental defectives and normal adults. *Amer. J. Psychiat.*, 1956, **112**, 803–807.
21. WILLIAMS, M. Psychophysiological responsiveness to psychological stress in early chronic schizophrenic reactions. *Psychosom. Med.*, 1953, **15**, 456–462.

Received July 6, 1956.

CONCLUSION-DRAWING, COMMUNICATOR CREDIBILITY, AND ANXIETY AS FACTORS IN OPINION CHANGE[1, 2]

BERNARD J. FINE

Boston University

HOVLAND AND MANDELL (8), investigating the relative effectiveness of communications whose facts and arguments are the same but whose issue-related conclusions are or are not explicitly stated, found an explicit statement of the conclusion more effective than no statement of the conclusion in stimulating opinion change. Thistlethwaite, de Haan and Kamenetzky (19) suggest that greater opinion change might be associated with the explicit communication primarily because it produces greater comprehension of the communicator's position. Although their explicit communication produced greater comprehension of the communicator's position, they found no difference in opinion change between their explicit and nonexplicit groups. However, a direct test of their implicit assumption of a positive relationship between knowledge of the conclusion and opinion change was not made. In view of the inconsistencies between the findings of these two studies, a further examination of the potential effectiveness of conclusion-drawing seemed warranted.

Hovland and Mandell also explored the interaction of source credibility with conclusion-drawing. Their predictions that a noncredible communicator would be more effective using a nonexplicit communication and that a credible communicator would be more effective with an explicit communication were not substantiated. However, they indicate that their induction of source credibility may not have been sufficiently potent. Accordingly, a second purpose of the current research was to test these predictions using a different induction procedure.

Since initial position (7, 11, 19) and audience anxiety (9, 10) have been shown to be relevant to opinion change, the interactive effects of both of these predispositional factors with conclusion-drawing were also examined. It was predicted that *Ss* initially opposed to the communicator's position would be more influenced by a nonexplicit communication whereas *Ss* initially favorable would be more influenced by an explicit one. With regard to anxiety, it was predicted that *Ss* high in anxiety would be more influenced by the anxiety allaying communication than would *Ss* low in anxiety. Furthermore, while it was felt that the explicit communication would be more effective than the nonexplicit one at both anxiety levels, it was predicted that the advantage of the explicit over the nonexplicit communication for *Ss* high in anxiety would be greater than for those low in anxiety.[3]

METHOD

Before Measures

The before questionnaire was presented to the *Ss* ostensibly as an opinion survey conducted by a national research organization. It contained four opinion items on biological warfare and a number of "filler" items on unrelated opinion issues. All items were multiple-choice in form. Also included were 20 items from the Taylor Manifest Anxiety Scale,[4] presented as a "personal inventory."

Experimental Communications and After Measures

Three weeks after completing the before questionnaire, the *Ss* read an article on biological warfare presented as part of a Boston University study investigating how well people learn from newspaper articles. The communication, following the "logical" form employed by Hovland and Mandell (8), took the position that biological warfare is not a "super" weapon and that we in the United States can defend against it. In

[1] The research is based on a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree at Boston University. The author is now at the Army Quartermaster Research and Development Center, Natick, Massachusetts. He wishes to thank Dr. Walter Weiss, his major professor, for his invaluable counsel in all phases of the study and Dr. Nathan Maccoby, his second reader, for his many contributions, especially those which made this study possible in the first place.

[2] The research was carried out under contract ONR 492(04) between the Office of Naval Research and Boston University. The program of research is under the general direction of Dr. Weiss.

[3] For a more detailed statement of the assumptions underlying these predictions, see (3).

[4] The twenty items were indicated by Harold P. Bechtoldt as being the most discriminating of all Taylor Scale items. Personal communication, July, 1955.

the explicit version, the final paragraph stated the conclusion that biological warfare is not a "super" weapon for us since we in the United States can defend against it. In the nonexplicit version, in place of the explicit statement of the conclusion, the *Ss* were urged to draw their own conclusions on the basis of the arguments presented in the article. The two communications differed in no way except for the last paragraph.

One half of the *Ss* read the explicit communication and one half read the nonexplicit. One half of the *Ss* in each of these groups had the authorship attributed to the New York Times, and the other half to the Daily Worker. The name of the source was mentioned both at the beginning and at the end of each article. For the Daily Worker, the explanatory phrase "official newspaper of the Communist Party in the United States" was stated at the beginning of the article. For the New York Times, the descriptive phrase "one of the most widely read newspapers in the U. S." was added. The four forms of the communication (conclusion-drawing × credibility) were distributed in accordance with a sequence of 4 x 4 Latin Squares so that no two adjacent *Ss* were assigned to the same experimental condition. The immediate-after questionnaire, in addition to containing the same four opinion items on biological warfare as the before questionnaire, also required the *Ss* to state the main position of the communicator and to make judgments concerning the article itself. Additional items measured acquisition of the content of the article, recall of source, and concern about biological warfare. Six weeks later the *Ss* were presented with the identical questionnaire ostensibly to "find out how well people remembered what they had read." After these questionnaires were collected, the *Ss* were informed of the true nature of the experiment.

Control Group

This group consisted of 36 students who were not exposed to the communications, but who completed appropriate questionnaires at the same times as the experimental groups. The control group was a single class used as a unit.

Subjects

Approximately 400 students, predominantly freshmen, at American International College, Springfield, Massachusetts, were used. The research was carried out in regularly scheduled class meetings in various psychology courses.[5]

Scoring

Each item on biological warfare was scored on a three-point scale. A response to an item indicating that biological warfare was not a "super" weapon was given a score of one; the "uncertain" category, a score of two; and response indicating that biological warfare was a "super" weapon, a score of three. The scores on the four opinion items were summated for each *S* for each administration, yielding a theoretical range from 4 to 12. To exhibit change of opinion from any one administration to another, changes in total scores between administrations were computed for each *S*, a plus change score indicating that the *S* had changed toward greater acceptance of the position that biological was a "super" weapon. The theoretical range of change scores was from minus eight to plus eight.

The anxiety items were scored by assigning a score of one to a response indicating that the *S* was anxious and a score of zero to a response indicating that he was not anxious. The *Ss* scoring in the upper *and* lower quartiles of the total distribution of anxiety scores were designated as high anxious and those in the middle two quartiles as low anxious.

In previous studies involving the Taylor scale, high scorers have typically been compared with low scorers, the comparison being based on the assumption that the extent of manifest anxiety as measured by the scale is directly and linearly correlated with the level of the internal anxiety drive (16, 17, 18). However, Eriksen (1, 2) has suggested that the Taylor scale may rather be an indicator of the ways in which individuals handle their anxiety, a view congruent with the lack of differential conditioning or learning effects found by some researchers between *Ss* scoring high and *Ss* scoring low on the Taylor scale (4, 6, 13, 14, 15). If a distinction is made between level of internal anxiety drive and the overt responses made to the cues from that drive, it thus seems reasonable to assume, following Eriksen, that high and low scores on the Taylor scale reflect differences in modes of response to anxiety rather than in level of drive, and that those scoring in the middle of the distribution derived from the Taylor scale are less anxious than those with high or low scores. This assumption is supported by the finding of a curvilinear relationship between learning and Taylor scores (14, 15), with *Ss* scoring high or low performing similarly to each other but differently from those scoring toward the middle of the distribution.

RESULTS

Table 1 shows the mean opinion change scores from before to immediate-after for the four experimental groups and the control group. All of the experimental groups showed a significantly greater amount of opinion change than did the control group ($p < .01$, two tails for each comparison). The experimental groups had not differed from the control group initially.

The total group given the explicit treatment showed a tendency toward greater opinion change than the nonexplicit group ($p = .10$, two tails by analysis of variance[6]). There was no significant difference between the credible

[5] The author wishes to thank Dr. Dorothy T. Spoerl and Dr. Charles A. Wells for the use of their classes. In addition, he is grateful to them and to Mr. Norman H. Berkowitz for assistance in administering certain phases of the research.

[6] Initial differences in opinion between the explicit and nonexplicit groups suggested an analysis of covariance. Such an analysis could not be undertaken, however, as a result of heterogeneity of regression between the groups.

TABLE 1

MEAN OPINION CHANGE SCORES FROM BEFORE TO IMMEDIATE-AFTER

Control Group		Experimental Groups			
		EC*	Ec	eC	ec
Means**	+.44	−2.26	−2.01	−1.68	−1.79
N	36	84	87	85	83

* E indicates explicit statement of conclusion, e nonexplicit; C indicates credible source, c noncredible.

** + indicates movement away from the communicator's position; − indicates movement toward the communicator's position.

and noncredible conditions in amount of opinion change from before to immediate-after.

The first item in the after questionnaire required the *S*s to state the communicator's position. In the explicit condition 62 per cent of the *S*s reported the communicator's position correctly, compared with only 40.5 per cent of the *S*s in the nonexplicit condition ($p < .001$, two tails). Incorrect responses to this question were mainly restatements of some of the general principles and specific considerations that the article suggested were necessary to arrive at a rational decision concerning the dangers of biological warfare.

To test the possibility that the difference in opinion change between the explicit and nonexplicit groups is due to the difference between them in *expressed* knowledge of the communicator's position, the relationship between *expressed* knowledge of the communicator's position and opinion change was examined. No significant difference in opinion change was found between those expressing knowledge of the communicator's position and those not expressing such knowledge.

Initial Position

The *S*s were divided at the median (6.47) into "favorable" and "unfavorable" groups in terms of their initial opinion scores. The mean opinion score for the favorables was 5.03; the mean for the unfavorables was 8.91. The unfavorable group showed greater opinion change than the favorable group ($p < .01$, two tails). The smaller opinion change of the favorable group is obviously related to the limited extent of change possible for those closer to the communicator's position. Within the favorable group the explicit treatment was more effective than the nonexplicit ($p = .02$, one tail).

TABLE 2

MEAN OPINION CHANGE SCORES FROM BEFORE TO IMMEDIATE-AFTER FOR *S*s INITIALLY HIGH OR LOW IN ANXIETY*

	EC	Ec	eC	ec	All Groups
Means for low anxious *S*s	−2.36	−1.74	−1.40	−1.38	−1.73
N	50	38	42	53	183
Means for high anxious *S*s	−2.12	−2.22	−1.95	−2.37	−2.15
N	34	49	43	30	156

* High anxiety *S*s are those scoring in either the upper or lower quartiles of the distribution of Taylor scale anxiety scores; low anxiety *S*s are those scoring in the middle two quartiles.

Within the unfavorable group there were no differences between the explicit and nonexplicit treatments.

Anxiety

The obtained range of scores on the Taylor scale was from 0 to 18 and the obtained mean was 6.04. While the distribution shows some skewness, this is not inconsistent with the distributions obtained by others using the scale on college populations (12, 22). The abridged Taylor scale had an odd-even reliability of .78 using the Rulon formula (5). Table 2 gives the mean opinion change scores from before to immediate-after for high and low anxious *S*s by experimental conditions. The total high anxious group shows a greater mean opinion change than does the total low anxious group ($p = .05$, one tail).[7]

The communication was equally effective for all conditions within the high anxious group; no opinion change differences were found between the explicit and nonexplicit and the credible and noncredible conditions. For low anxious *S*s, however, the explicit treatment was more effective than the nonexplicit ($p = .03$, one tail).

Acquisition of Content

To check the comparability of the groups with respect to acquisition of content of the communication, fact-quiz scores for the experimental groups were compared. (The fact-

[7] The separate mean change scores for the upper and lower quartiles of the anxiety distribution were greater than that for the combined middle two quartiles. However, only the *S*s in the upper quartile changed significantly more than those scoring in the middle two quartiles.

quiz consisted of six multiple-choice items; the mean for all *Ss* was 5.09). The *Ss* reading the explicit communication show a significantly higher mean fact-quiz score than do the *Ss* reading the nonexplicit communication ($p = .06$, two tails). However, no relationship between acquisition of content (fact-quiz) and opinion change was found.

Judgmental Items

There were no significant differences between the credible and noncredible or the explicit and nonexplicit groups in the proportion of *Ss* saying that the article was fair or a piece of propaganda. Responses to both of the relevant judgmental items indicated that approximately 70 per cent of all the *Ss* felt that the article presented a fair treatment of the topic and was not propagandistic in intent.

Recall of Source

An open-ended question asking the *S* to write the name of the source was included in the after questionnaire. Table 3 shows the percentage of *Ss* in each experimental group recalling the source correctly. The *Ss* in the credible condition were more likely to recall the source correctly than were the *Ss* in the noncredible condition ($p < .02$, two tails). There was no significant difference between the explicit and nonexplicit groups in the proportion of *Ss* recalling the source correctly. No relationship was found between opinion change and recall of source.

Concern About Biological Warfare

After reading the communication, the *Ss* were asked two questions regarding their concern about biological warfare ("How concerned are you about biological warfare?") and the importance of the topic to them ("How important is the topic of biological warfare to you?"). Because the responses to the two questions were very highly correlated, the *Ss* were divided into two groups; those that had responded that they were "extremely" or "very" concerned about biological warfare *and* who also felt that the topic was "extremely" or "very" important to them were designated as highly concerned and all other *Ss* as relatively nonconcerned. Opinion changes between and within the two groups are shown in Table 4. The *Ss* who indicated relatively low concern about biological warfare, after reading the communication, showed a greater mean opinion change than those who expressed high concern ($p < .01$, two tails). The two groups did not differ in initial opinions.

TABLE 3

PERCENTAGE OF *Ss* RECALLING SOURCE CORRECTLY

	Experimental Groups			
	EC	Ec	eC	ec
Percentage	78.5	64.4	76.5	69.9
N	84	87	85	83

TABLE 4

MEAN OPINION CHANGE SCORES FROM BEFORE TO IMMEDIATE-AFTER FOR *Ss* RELATIVELY HIGH AND LOW IN CONCERN ABOUT BIOLOGICAL WARFARE

Degree of Concern	Experimental Group				
	EC	Ec	eC	ec	All Groups
Highly Concerned	− .64	−1.46	− .82	−1.67	−1.20
(*N*)	(11)	(13)	(22)	(24)	(70)
Not highly Concerned	−2.51	−2.11	−1.98	−1.76	−2.11
(*N*)	(73)	(74)	(63)	(59)	(269)

The *Ss* in the nonexplicit treatment were more likely than the *Ss* in the explicit treatment to say that biological warfare was of concern and importance to them ($p < .01$, two tails).

Delayed-After Results

After six weeks, the mean opinion score of each experimental subgroup mentioned in the analysis of the immediate-after data was still significantly different from the mean of the control group ($p < .01$, two tails). With respect to the comparisons between experimental subgroups, only the difference between the high and low concern groups approaches significance ($p = .11$, two tails). The latter group showed greater retained opinion change than the former.

At this time, the *Ss* in the credible condition were more likely than those in the noncredible condition to say that the author gave a fair presentation ($p < .05$, two tails); and there was a greater tendency for the *Ss* who read the nonexplicit communication than for those who read the explicit one to indicate that the author's treatment of the topic was fair ($p < .10$, two tails).

DISCUSSION

The indication that the explicit treatment is somewhat more effective than the nonexplicit tends to support the findings of Hovland and Mandell (8). However, the basis for the potentially greater effectiveness is not certain. Under certain conditions, knowledge of the communicator's position may be an important factor as Thistlethwaite, de Haan and Kamenetzky (19) suggest. In the present research, however, no relationship was found between *expressed* knowledge of the communicator's position and opinion change. This finding is consistent with a conceptualization of opinion change as a continuous process in which the *S*, while reading the communication, is being affected by many content cues. While some people may not be able to order verbally everything they have read in a communication in a fully organized, meaningful, and closed form, they are apparently able to get enough of the "drift" of the communication to be as influenced by it as are others who can explicitly express the communicator's position.

To account for the superiority of the explicit condition over the nonexplicit in changing opinions, different assumptions are needed. Given a fairly complex communication, it is quite likely that the person, while reading, does not adequately perceive the logical organization of the content or arrange the data and arguments to form the implied syllogistic pattern represented by the current communication. The more complex the communication, the more likely he will have difficulty in grasping this informative patterning. Reading the conclusion may instigate awareness of the relationship existing between the preceding factual and logical groundwork and the basic point of the message. Thus, the explicit presentation of the conclusion may serve essentially to unify the body of the communication so that it becomes more clearly related to the conclusion. The impact of this unified, better understood communication is likely to be greater than in the nonexplicit condition where the person, if he has inferred the desired conclusion, may well be less certain of its appropriateness and validity. While it is probable that the person who arrives at the appropriate conclusion on his own achieves a degree of unification in the content of the message, the attendant uncertainty should serve to reduce the over-all impact of the communication. Unfortunately, no measure of certainty was included in this study.

If the recipient of a communication about a potentially anxiety arousing topic is uncertain concerning its anxiety allaying conclusion, however, somewhat greater residual concern would be expected than if he were more certain about it. In line with this expectation, the data show that *S*s in the nonexplicit condition evidenced more concern than the *S*s in the explicit condition. If uncertainty of the correct conclusion underlies residual concern, then less opinion change may be expected among the highly concerned *S*s than among those not as concerned. Again, the data bear this out (see Table 4).

The lack of opinion change differences between the credible and noncredible conditions may have been due to a weakness in the induction of the variable, but this possibility seems unlikely. It had not been considered necessary to obtain measures of the credibility attributed to the two sources by the *S*s, since previous experience with other *S*s indicated that the New York Times and the Daily Worker were perceived as differing greatly in credibility. To heighten this difference, qualifying phrases amplifying the credibility or lack of credibility were attached to the names of sources, it will be remembered, in the present study.

The results for initial position were possibly complicated by the generally favorable initial opinions of *all* the *S*s. This fact may have additional consequences if it is assumed, as seems reasonable, that persons form an accurate impression of a communicator's position more readily when it is discrepant from their own. Here, even in the nonexplicit condition, *S*s may be able to sense the main tenor of the communication because its message is so strongly opposed to their own position. But when the position of the *S* is not very discrepant from the communicator's, a finer discrimination is required. Thus, it might be expected that an explicit statement of the appropriate conclusion would result in greater opinion change for the initially favorable *S*s than would no explicit statement.

The finding that *S*s high in anxiety were more influenced by the communication than were those low in anxiety is consistent with the results of previous research (20, 21) in which

congruency between relevant response predispositions and the opinion response urged by the communicator were found to facilitate communication effectiveness.

SUMMARY

The study was designed to explore the simple and interactive effects of conclusion-drawing, communicator credibility, initial position, and anxiety. A before-after design was employed. The *Ss*' initial positions on biological warfare were assessed and they were given the Taylor Manifest Anxiety Scale. Three weeks later all *Ss* read an article on biological warfare which took the position that biological warfare was not a "super" weapon. This was followed immediately by another opinion questionnaire containing opinion items, items testing retention of content of the communication and items pertaining to the *Ss*' judgments of various aspects of the communication. Six weeks later all *Ss* took identical questionnaires. A control group that was not exposed to the experimental communication but took all of the questionnaires was employed. Some of the major findings are: (*a*) No relationship was found between *expressed* knowledge of the communicator's position and opinion change; (*b*) the explicit treatment was somewhat more effective in changing opinions than was the nonexplicit; (*c*) no difference was found in opinion change between the groups given credible and noncredible treatments; (*d*) the opinion change of *Ss* high in inferred anxiety was greater than those low in anxiety; and (*e*) *Ss* high in residual concern changed less in opinion than *Ss* not high in residual concern.

REFERENCES

1. ERIKSEN, C. W. Some personality correlates of stimulus generalization under stress. *J. abnorm. soc. Psychol.*, 1954, **49**, 561–566.
2. ERIKSEN, C. W., & DAVIDS, A. The meaning and clinical validity of the Taylor Anxiety Scale and the Hysteria-Psychasthenia Scale from the MMPI. *J. abnorm. soc. Psychol.*, 1955, **50**, 135–138.
3. FINE, B. J. Conclusion-drawing, communicator credibility and anxiety as factors in opinion change. Unpublished doctor's dissertation, Boston Univer., 1956.
4. FRANKS, C. M. Conditioning and personality. *J. abnorm. soc. Psychol.*, 1956, **52**, 143–151.
5. GUILFORD, J. P. *Psychometric methods* (2nd ed.). New York: McGraw-Hill, 1954.
6. HILGARD, E. R., JONES, L. V., & KAPLAN, S. J. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, **42**, 94–99.
7. HOVLAND, C. I., LUMSDAINE, A. A., & SHEFFIELD, F. D. *Experiments on mass communication*. New Jersey: Princeton Univer. Press, 1949.
8. HOVLAND, C. I., & MANDELL, W. An experimental comparison of conclusion-drawing by the communicator and by the audience. *J. abnorm. soc. Psychol.*, 1952, **47**, 581–589.
9. JANIS, I. L. Personality correlates of susceptibility to persuasion. *J. Pers.*, 1954, **22**, 504–518.
10. JANIS, I. L. Anxiety indices related to susceptibility to persuasion. *J. abnorm. soc. Psychol.*, 1955, **51**, 663–668.
11. JANIS, I. L., LUMSDAINE, A. A., & GLADSTONE, A. I. Effects of preparatory communications on reactions to a subsequent news event. *Publ. Opin. quart.*, 1951, **15**, 487–518.
12. KERRICK, JEAN S. Some correlates of the Taylor Manifest Anxiety Scale. *J. abnorm. soc. Psychol.*, 1955, **50**, 75–78.
13. KERRICK, JEAN S. The effects of manifest anxiety and IQ on discrimination. *J. abnorm. soc. Psychol.*, 1956, **52**, 136–138.
14. MATARAZZO, RUTH G., & MATARAZZO, J. D. Anxiety level and pursuitmeter performance. *J. consult. Psychol.*, 1956, **20**, 70.
15. MATARAZZO, J. D., ULETT, G. A., GUZE, S. B., & SASLOW, G. The relationship between anxiety level and several measures of intelligence. *J. consult. Psychol.*, 1954, **18**, 201–206.
16. SPENCE, K. W., & TAYLOR, JANET A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.
17. TAYLOR, JANET A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–89.
18. TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.
19. THISTLETHWAITE, D. L., DE HAAN, H., & KAMENETZKY, J. The effects of "directive" and "nondirective" communication procedures on attitudes. *J. abnorm. soc. Psychol.*, 1955, **51**, 114–119.
20. WEISS, W., & FINE, B. J. Opinion change as a function of some intra-personal attributes of the communicatees. *J. abnorm. soc. Psychol.*, 1955, **51**, 246–254.
21. WEISS, W., & FINE, B. J. The effect of induced aggressiveness on opinion change. *J. abnorm. soc. Psychol.*, 1956, **52**, 109–115.
22. WOLFF, W. M. Certainty: generality and relation to manifest anxiety. *J. abnorm. soc. Psychol.*, 1955, **50**, 59–65.

Received July 3, 1956.

THE ROLE OF EXPERIENCE IN JUDGING CHILDREN'S PHOTOGRAPHS

WILLIAM KESSEN[1]

Yale University

In a recent review of the psychological literature on the perception of people, Bruner and Tagiuri conclude: "The first step in reacting to another is forming an impression of him. Later reactions depend on this first step. If there is to be a science of interpersonal behavior, it will rest upon a cornerstone of social perception" (1, p. 650). This assertion is of special practical and theoretical importance when the perception of children is under consideration. For one, whatever impressions a parent forms of his child will affect, in as yet undetermined ways, subsequent interactions between parent and child, and it is important to determine to what degree variations in the perception of children can be ascribed to the stimulus presented or, failing that, to determinants in the personality of the parent. An approach to this problem can be made by evaluating the *consistency* with which different parents judge the same child stimulus. Moreover, by comparing with this evaluation some independent index of *validity* there is the additional gain of determining whether parents, when compared with nonparents, have a greater or lesser accuracy of judgment as a result of their intimate contact with at least one member of the class of child stimuli.

Problems of much the same order can be stated for the professional student of human behavior, the psychologist. Does his training or clinical experience modify either the consistency or the validity of his perceptions of people? The judgment of children, particularly in the younger range, presents a case of special importance, because many of the cues for personality evaluation available from adult human beings are not present, perhaps forcing the psychologist back to what some writers have called "physiognomic perception" (5, 6).

The present study was designed as an initial approach to this area of problems by comparing the consistency and validity of judgments made of children's photographs, using judges who varied in their experience with, or knowledge of, children.

METHOD

From the photographic files of the Yale Child Study Center longitudinal study (2, 7), glossy photographs of 10 children (6 girls and 4 boys) were selected. At the time the pictures were taken, the children ranged in age from 14 to 23 months, the median age being 19 months. The pictures were full length and the children were naked. When several photographs of a single child were available, selection was made on the criteria of (*a*) being as near full-face to the camera as possible, and (*b*) reduction of extraneous objects (toys, mothers' hands, etc.). The 10 photographs selected were reprinted to a common size (4″ x 8″), mounted on heavy white cardboard, 8″ x 10″, and ordered alphabetically by the code names of the children.

This series of photographs was presented to 50 *Ss* in individual sessions. The *Ss* fell into three groups:

1. Undergraduates: Twenty male undergraduates, all childless.

2. Parents: Ten junior faculty members and their wives, who met the additional requirements that none was a psychologist, and that they had a child at least as old as those shown in the photographs, but no more than one year older.

3. Experts: Ten professional psychologists, either faculty members or senior graduate students. Three of these also met the "relevant child" requirement imposed on the parent group.

The instructions given to all *Ss* were as follows:

I'm interested in how well people are able to tell something about children on the basis of a single photograph. There are ten pictures in the series, of six girls and four boys, just past their first birthday. What I want you to do is rank them, that is, put them in order, for each of several traits.

For example, how would you rank these pictures as showing *intelligence*—how bright the child is.

Now, would you put them in order again, this time as a function of *sociability*—how well they react to and interact with people.

The next characteristic I'd like you to sort on is *activity*—how vigorous and active the child is.

And now, could you rank the pictures on *sensitivity*—that is, how readily the child responds to *things* going on around him or her.

And finally, I'd like you to judge *mental health*—the order of good functioning or healthy adjustment.[2]

[1] The writer is indebted to Professor Katherine M. Wolf who suggested the research and made the "contemporary validity" judgments, and to Dr. Sally Provence, who made the "predictive validity" judgments.

[2] All *Ss* were also asked to select the most attractive and least attractive boy and girl. Because an ordering

At the end of each sort, *E* recorded the *S*s' ranks and returned the pictures to their original order before presenting them again. No special restrictions were put on methods of handling the mounted photographs, but almost all *S*s spread the pictures out and then indicated their rank order by seriating the pictures spatially or restacking them in the chosen order. If any *S* made comments or asked questions about the instructions, *E* made some nondirective comment, or accepted without comment the definition the *S* set for the task. In some cases when the *S* persisted in questioning the task (almost exclusively for the Sensitivity ranking), *E* gave a paraphrase of the original instructions.

As an index of validity, a child psychologist who had observed the children intensively through their first two years, made rankings on the same instructions, with a further request to make her judgments relevant to the first half of the children's second year. She was not shown the photographs but rather small white cards on which the code names of the children were typed.

To obtain a "prediction" index, a pediatrician of the longitudinal study also ranked the cards on the instructions given earlier, but with the qualification that her judgments be on the basis of the last time she had seen the child. Because of attrition in the study sample, this last contact ranged from age 16 months to age 5 years, with the median at age three and one-half years.

Beyond these judgmental indexes of validity, Hetzer-Wolf (3) test results were available for all children as a validity check of the Intelligence ranking.

RESULTS

Acceptance of the Task

A majority of *S*s in all groups met the initial instructions with the remark: "Impossible" or a similar reaction of doubt, and then proceeded with scarcely a pause to perform the task. No questions were asked about the Intelligence (*I*), Activity (*A*) or Health (*H*) instructions, and only two about the Sociability (*Soc*) ranking (both from Experts working on socialization). The Sensitivity (*Sen*) ranking caused a fairly large number of hesitations and questions, however, turning in a more or less sophisticated way on the issue of whether the instructions asked for *quality* of sensitivity (does the child respond adequately?) or for *thresholds* (is the child emotional, irritable?). This confusion is reflected in the results reported below.

Concordance among Ss.

The first questions put to the data were: How consistent are the *S*s' responses, and are there variations in consistency assignable to groups? The index chosen was Kendall's coefficient of concordance *W* (4), and the values obtained are shown in Table 1. This coefficient indicates the degree to which *S*s make the same rankings. Twelve of the 15 independent statistics shown in the upper lines of Table 1 are significant, with those for *A* and *Sen* being consistently lower than the other three rankings for all groups. Moreover, there seems to be no systematic variation among the groups, the Undergraduates being as consistent as either of the experienced groups. The variation in consistency shown, therefore, is assignable to differences among scales rather than among groups of *S*s.

The degree of consistency shown in the upper lines of Table 1 was based on groups of 20, 20 and 10 judges. If, however, each group is seen as representing a single judge (by calculating mean ranks for each child for each group) and *W* recomputed on this basis, the values increase markedly in size (Table 1). The same comparisons are represented in the more usual form of correlation coefficients (Kendall's *tau*) in Table 2, showing the cor-

TABLE 1

CONCORDANCE (*W*) WITHIN GROUPS OF JUDGES (In the Combined Case, Each Group is Treated as a Single Judge)

Group	Variable Ranked					
	N	*I*	*Soc*	*A*	*Sen*	*H*
Undergraduates	20	.31***	.44***	.15**	.08	.44***
Parents	20	.23***	.48***	.18***	.06	.35***
Fathers	10	.26**	.48***	.12	.23*	.35***
Mothers	10	.26**	.52***	.33***	.08	.46***
Experts	10	.38***	.39***	.07	.27**	.39***
Combined	50	.92**	.90**	.85**	.65*	.88**

* $p < .05$.
** $p < .01$.
*** $p < .001$.

TABLE 2

CORRELATIONS (*TAU*) AMONG GROUPS (Each Group is Represented by Mean Values)

Group	Variable Ranked				
	I	*Soc*	*A*	*Sen*	*H*
Parents-Underg.	.73**	.67**	.66**	.54*	.78***
Parents-Experts	.63**	.72**	.72**	.39	.64**
Underg.-Experts	.85***	.78***	.55*	.18	.60**

* $p < .05$.
** $p < .01$.
*** $p < .001$.

procedure different from that used with rankings was necessary to evaluate this rating, no results drawn from it are included in this report.

relation of the mean ranks of each group with those of the other groups.

These results indicate that, at least for the present task, a high level of consistency is obtained when each group is considered a single judge and compared with other groups, and a lower but significant level of consistency is obtained when individual judges are compared within groups. Moreover, the three groups do not vary in their consistency of judgment and are making, to a significant degree, the *same* rankings.

The Parent group can be dichotomized by sex to evaluate consistency variations on this dimension. Table 1 presents the obtained *W*'s, and except for the paradoxical finding that fathers are more consistent in the rankings of Sensitivity and mothers more consistent in the ranking of *A*, there are no large overall differences between male and female parents. Incidentally, it was found that husband-wife pairs do not show higher correlations of judgment than do unrelated men and women.

Halo Effects.

For each *S*, intercorrelations of each ranking with every other ranking were made, to determine the tendency of *S*s to make the same ranking of the pictures regardless of instruction, i.e., to assess halo effect. The 10 obtained values for each *S* were averaged without regard to sign. Table 3 summarizes these calculations, giving range and median *tau* value for each group on each ranking. As can be seen, the range of variation among groups is narrow, with only a slight difference separating the Undergraduates from the experienced groups. In effect, there is no consistent tendency for halo effect to vary among groups of judges.

Validity of Judgments.

Each ranking of the 50 judges on the five scales was correlated with the rankings made by the child psychologist who had studied the children intensively ("contemporary validity"). The results of these computations are shown in Table 4 which lists the median value within each group of the correlation coefficients (again, Kendall's *tau*). They are in general low, and no median value is significantly different from zero. If, however, a validity run analogous to the consistency run of Table 2 is made, that is, considering each group to be a single judge, much higher coefficients are obtained (Table 5). Not only are the rankings consistent when combined, but they also become clearly more valid. Of the 15 independent statistics, 9 are significantly different from zero and in the proper direction, and two (Undergraduates-*Soc* and Experts-*H*) are barely short of the five per cent level. The *Sen* index is once more deviant, even showing one significant correlation of negative sign. However the comparisons are made, there are no gross differences among groups. If the *Sen* correlations are withdrawn as belonging somehow to a different population, there is a slight tendency for the mean correlations (*tau*) from Table 5 to favor the Parents (.52) over the Undergraduates (.46) and the Experts (.54) over the Parents.

The correlations with the validity criterion of the male parents considered as a single judge and the female parents so considered are also shown in Table 5, a further demonstration that differences on this dimension are small and apparently unsystematic.

To appraise "predictive validity" the rankings of the groups considered as three single judges were correlated with the rankings of the staff pediatrician who had examined the children throughout their stay in the longitudinal research, and who made her judgments relevant to the last time the children were seen. Table 6 shows that the relation-

TABLE 3

MEDIAN INTERCORRELATION (*TAU*), AND RANGE OF INTERCORRELATION FOR EACH GROUP, OF ALL SCALES WITHIN SUBJECTS

(i.e., the Degree of Halo Effect)

Group	Median *tau*	Range for Group
Undergraduates	.41	.16–.65
Parents	.34	.17–.63
Experts	.32	.20–.72

TABLE 4

MEDIAN VALIDITY CORRELATIONS

(Each *S*'s Rankings are Correlated with the Criterion)

Group	Variable Ranked				
	I	*Soc*	*A*	*Sen*	*H*
Undergraduates	.39	.29	.09	−.11	.24
Parents	.22	.38	.29	−.11	.24
Experts	.24	.38	.13	−.07	.29

TABLE 5
CORRELATIONS (*TAU*) WITH VALIDITY CRITERION
(Each Group Considered as a Single Judge)

Group	Variable Ranked				
	I	*Soc*	*A*	*Sen*	*H*
Undergraduates	.56*	.42	.39	−.33	.47*
Parents	.56*	.49*	.56*	−.54*	.47*
Fathers	.38	.57*	.42	−.36	.47*
Mothers	.54*	.49*	.60**	−.20	.38
Experts	.54*	.69**	.49*	−.27	.42

* $p < .05$.
** $p < .01$.

TABLE 6
CORRELATIONS (*TAU*) WITH PREDICTION CRITERION OF VALIDITY
(Each Group Considered as a Single Judge)

Group	Variable Ranked				
	I	*Soc*	*A*	*Sen*	*H*
Undergraduates	.24	.47*	−.11	−.14	.47*
Parents	−.07	.22	−.02	−.40	.33
Experts	.18	.51*	.04	−.45	.29

* $p < .05$.

ships decay to a point just short of chance expectancy, even though the correlation (*tau*) between the two validity judges is marked and consistent (*I*—.42, *Soc*—.47, *Sen*—.60, *H*—.73, all values but the first significant at the .05 level or better). In a word, the validity of judgment by the *S*s is higher for "contemporary" judgments than for "predictive" ones, although *Soc* and *H* correlations in the latter case are sometimes significant.

A final index of validity is provided by the correlations of the *I* rankings of the groups, considered as three single judges, with scores on Hetzer-Wolf (3) tests given within six months of the time each child was photographed. For the Undergraduates and the Parents, the correlation was .47 and for the Experts, it was .54, values significant at the .05 level. The comparable values for concontemporary validity were .56, .56 and .54 (Table 5). This finding is particularly important in demonstrating that the validity judge did not merely guess accurately at the expected rankings of the *S*s.

DISCUSSION

Among the findings obtained in this study, those which warrant some special mention are (*a*) the high consistency and validity of groups as judges compared with individuals as judges, and (*b*) the failure to show variations in consistency or validity on the dimension of experience with children, practical or professional.

The first problem seems to be answered in a rather straightforward way. By pooling judgments there is a tendency for idiosyncratic variations to be eliminated, much as the probability of a particular die face appearing is more stably measured with an increasing number of rolls. One source of such idiosyncratic responding (probably not random) in this study was expressed explicitly by a number of parents who remarked on the similarity of one or another of the photographs to their own children. Presumably other such biases, different in different subjects, operated to introduce error into the rankings of individual judges.

This line of discussion leads to a more difficult problem—the assumption that there is something stable in the situation for the errors to distort, something analogous to a particular die face. Nothing in the results presented here contradicts such an assumption, although the correlations of *S*s' rankings with the validity rankings are modest, even when statistically significant.[3] To get at this something which is perceived by the *S*s, one can make two attacks, either a systematic variation in cues presented (changes in the photographs; for example, restriction to face alone), or an examination of response differences among individuals or groups. This second goal might be achieved by interviewing "good" and "bad" judges, or by comparing groups of judges who differ in some definable way. It was the latter procedure which was followed in the present study. The results of comparing groups selected to vary in their experience with children were disappointing insofar as they showed no systematic differences in any characteristic (concordance, halo, validity) which could be

[3] This statement is mitigated somewhat by the consideration that the photographs used were not, like those in many social perception studies, posed; the children were in more ways than one *au naturel*. However, the fact that the validity correlations are substantially lower than the consistency correlations points to the danger of substituting implicitly the one for the other in social perception research. Agreement does not necessarily indicate accuracy.

assigned to group membership. It seems therefore that this first step in the analysis of the perception of children has produced a weak conclusion—that the characteristics measured do not vary with experience of the judges with the class of objects judged.

SUMMARY AND CONCLUSIONS

Still photographs of ten children in their second year were presented to 3 groups of *Ss* (20 undergraduate men, 20 parents and 10 professional psychologists) in individual sessions. All *Ss* were instructed to rank the photographs on Intelligence, Sociability, Activity, Sensitivity, and Mental Health. Validity of the rankings was indicated by correlation with the rankings of a child psychologist who had observed the children closely throughout their first two years.

The results obtained appear to warrant the following conclusions:

1. All groups of *Ss* made consistent judgments of Intelligence, Sociability, and Health. Concordance measures for Activity and Sensitivity were lower for all groups.

2. This consistency did not vary with experience with children, defined either on a practical (Parents) or on a professional training (Experts) basis.

3. For the present task, groups of *Ss* combined to provide a single measure were highly consistent with other such combined groups.

4. Median validity correlations of individual *Ss* on all rankings were low, and not statistically significant.

5. When the groups were combined to function as single judges, validity on four of the rankings (all but Sensitivity) was high and, in nine of twelve correlations, statistically significant.

6. There were no striking differences in validity, however determined, among the three groups, or between male and female parents.

7. These findings suggest that whatever perceptual variables underlie judgments of children's photographs, these variables are not demonstrated to be related to practical experience or to professional training.

REFERENCES

1. BRUNER, J. S., & TAGIURI, R. The perception of people. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge, Mass.: Addison Wesley, 1954.
2. COLEMAN, R. W., KRIS, E. & PROVENCE, S. The study of variations of early parental attitudes. *Psychoanal. Stud. Child*, 1953, **8**, 20–47.
3. HETZER, H., & WOLF, KATHERINE. Babytests. *Z. f. Psychol.*, 1928, **107**, 62–104.
4. KENDALL, M. G. *Rank correlation methods*. London: Griffin, 1948.
5. KOFFKA, K. *Principles of Gestalt psychology*. New York: Harcourt Brace, 1935.
6. WERNER, H. *Comparative psychology of mental development*. Chicago: Follett, 1948.
7. WOLF, KATHERINE M. Observation of individual tendencies in the first year of life. In M. J. E. Senn (Ed.), *Problems of infancy and childhood*. New York: Macy, 1953.

Received July 9, 1956.

RELATIONSHIPS BETWEEN CHILD TRAINING AND THE PICTORIAL ARTS[1]

HERBERT BARRY, III

Yale University

An artist's personality is commonly thought to be expressed in the style of his art creations, so that individual differences in art style are related to individual differences in personality. If this assumption is true, it should be possible to find a correlation between an appropriate index of personality and art style. The index of the artist's personality may include events known to be influential, such as the socialization practices he has undergone. A correlation between socialization and art style, if obtained, may mean that the socialization and art variables are related to the same aspect of the artist's personality.

This argument may be generalized from the single individual to the typical or modal individual in a small and relatively homogeneous society. The socialization practices and art styles as aspects of culture are perpetuated beyond the individual's life span by custom, but cultural custom like an individual's behavior develops and may be modified as an adjustment to the personality characteristics of the individuals. The present study was carried out to test for a correlation between severity of socialization and style of art among a sample of nonliterate societies. The measures on socialization have been presented by Whiting and Child (4), and the present investigator independently obtained measures of art style.

Method

Thirty nonliterate societies were selected from the list of 76 on which Whiting and Child gave socialization data. These 30 societies comprised all those from which the investigator was able to find at least ten works of graphic art, either exhibited in museums, or as illustrations in ethnographic reports. If 10–20 works of art were available for a culture they were all rated; if more than 20 were available, representative ones were picked at random. In all, 549 works of art were rated.

Each work of art was rated on 18 criteria of art style, using a 7-point scale. Since the score on most of the variables was influenced by the material used, the art works of each culture were compared with those of other cultures in the same material, e.g., weaving, woodwork, wickerwork, and pottery. The culture was designated as above or below the median in each variable, combining its score in the different materials but comparing its art in each material only with the art of the other cultures in the same material.

Whiting and Child described five universal systems of behavior: oral, anal, sexual, dependence, and aggression. Two stages of socialization were distinguished for each system: initial satisfaction and socialization anxiety. The resultant 10 variables had each been rated on a 1–7 scale by three judges, and the three ratings added together. This combined score for each of the 10 variables was transformed into standard scores, using the sample of 30 cultures on which art scores were available. The 5 scores of initial satisfaction were reversed in sign so that a high score would mean low satisfaction. The 10 standard scores were then combined for an overall measure of severity of socialization, with a high score signifying high severity.

Results

Eleven art variables were considered to be measures of complexity of art style.[2] One of these variables, entitled complexity of design, was defined at the upper extreme as a design with many unrepeated figures to form a complex organization of design; the lower extreme of this variable was defined as a design with few figures or repetition of figures to form a simple organization of design. The other 10 variables, which were less directly defined in terms of complexity of art style, are listed in Table 1.

To provide a combined measure of complexity of art, cultures were scored as high in combined complexity if they were rated above the median (in the direction of complexity) in 6 or more of the 11 separate measures, low if they were above the median in 5 or fewer of these measures. Table 1 (left-hand column of figures) shows the tetrachoric correlation of

[1] This paper is a revised version of a thesis presented to the Social Relations Department at Harvard College in partial fulfilment of the requirements for the B.A. degree with honors. The thesis was written under the generous and helpful guidance of Dr. John W. M. Whiting. The revisions were made with helpful suggestions from Dr. Irvin L. Child.

[2] Seven additional variables were rated but are omitted here for the following reasons: 3 concerned color and there were insufficient cases; 2 were unreliably rated; 2 were rated with insufficient variation.

TABLE 1

CORRELATION OF THE ART VARIABLES WITH COMPLEXITY OF STYLE AND SEVERITY OF SOCIALIZATION

(All of the correlations are positive.)

Art Variables	Complexity of Style (r_{tet}, $N = 30$)	Severity of Socialization (r_{bis}, $N = 28$)
Complexity of design	.98**	.71**
Presence of enclosed figures	.91**	.32
Presence of lines oblique to each other	.91**	.13
Presence of sharp figures	.81**	.18
Presence of curved lines	.81**	.07
Representativeness of design	.67*	.56*
Presence of lines oblique to edges	.67*	.45
Crowdedness of space	.67*	.12
Asymmetry of design	.50	.26
Presence of border	.31	.20
Shortness of lines	.31	.11

* $p < .05$.
** $p < .01$.

each variable with this combined measure, using the table compiled by Davidoff and Goheen (1). The statistical significance was computed by the Fisher-Yates exact test (2).

The biserial correlation of each variable of art style with the overall measure of severity of socialization is shown in the right-hand column of Table 1. For each of the 11 art variables, the direction of high complexity is positively related to high severity of socialization. Two of the art variables are significantly related to the overall measure of socialization by t test. Complexity of design, which is most closely related to the combined measure of complexity of art, is the art variable showing the highest correlation with severity of socialization. The combined measure of complexity of art has a biserial correlation of +.47 with the overall measure of severity of socialization ($t = 2.06$; $p < .05$).

The biserial correlation of complexity of design with each of the ten measures of socialization is shown in Table 2. In all 10 cases high severity of socialization is positively correlated with high complexity of design, although the intercorrelations among the 5 systems of behavior (reported by Whiting and Child) are mostly slight.

Table 3 lists the 30 cultures in this sample, grouped into those above and below the median on complexity of design. The left-hand column of figures for both groups of cultures shows the combined measure of severity of socialization. The right-hand column indicates that complexity of design is correlated with the occurrence of cultures for which some of the socialization ratings were omitted because of insufficient information. Ratings were omitted in none of the 15 cultures above the median in complexity and in 7 of the 15 cultures below the median. This difference is significant by the Fisher-Yates Exact Test ($p < .01$).

Instances of severe socialization practices are generally more dramatic and likely to be reported more fully. Therefore it is probable

TABLE 2

CORRELATION OF COMPLEXITY OF DESIGN WITH MEASURES OF SEVERITY OF SOCIALIZATION

(Satisfaction and anxiety measures taken from Whiting and Child (4))

Behavior system	Satisfaction		Anxiety	
	r_{bis}	N	r_{bis}	N
Oral	−.59**	29	+.35	27
Anal	−.32	23	+.16	23
Sexual	−.05	28	+.28	28
Dependence	−.46	27	+.42	26
Aggression	−.04	30	+.33	30

** $p < .01$.

TABLE 3

COMPLEXITY OF DESIGN AND SEVERITY OF SOCIALIZATION IN THE SAMPLE CULTURES

Complexity of Design Above Median			Complexity of Design Below Median		
Culture	Socializ. severity	Not rated	Culture	Socializ. severity	Not rated
Ashanti	+1.01	0	Hopi	+.37	0
Chiricahua	+.77	0	Thonga	+.34	0
Dahomean	+.69	0	Navaho	+.20	0
Alorese	+.38	0	Paiute	−.05	0
Western Apache	+.35	0	Masai	−.28	4
Kwakiutl	+.29	0	Papago	−.30	0
Samoans	+.26	0	Ifugao	−.45	2
Ainu	+.12	0	Andamanese	−.54	2
Maori	+.12	0	Marshallese	−.56	3
Marquesan	+.09	0	Chenchu	−.58	0
Arapesh	+.06	0	Yagua	−.62	0
Yakut	−.01	0	Comanche	−.94	0
Balinese	−.21	0	Murngin	−1.05	4
Trobrianders	−.22	0	Omaha	—	6
Teton Dakota	−.47	0	Zuni	—	8

Note:—Cultures are divided into above and below the median for the art variable of complexity of design. Each culture is listed with its average score on the combined measure of severity of socialization, and the number of socialization variables (out of 10) on which a rating was not made.

that omissions of ratings indicate an unreported low degree of severity more often than an unreported high degree of severity. There is evidence for this inference from the fact that the socialization ratings made with the lower of two specified degrees of confidence were generally below average in severity. If the majority of omissions indicate low severity of socialization, they give further evidence of the positive relationship described above with low complexity of art design. Since the art materials and socialization information were obtained from different sources, it is unlikely that the correlation indicates a direct causative relationship between inadequate ethnographic information and availability of art works below the median in complexity of design.

DISCUSSION

The correlation between severity of socialization and complexity of design in art works indicates the presence of a connecting link between these two variables, to which both are related. Cultures with low and high severity of socialization are certain to differ from each other in the typical personality characteristics of their people, if we believe that socialization has an important influence on personality and that our measures of socialization are valid. Therefore, cultures with simple and complex art style also differ from each other in the same typical personality characteristics of their people. It is reasonable to believe that these personality characteristics influence the style of the art works created by the people. This belief provides a possible explanation for the correlations between socialization and art variables.

The correlations of the 11 variables of art style with the overall measure of art style suggest the presence of a general factor of complexity or elaborateness in various specific measures of style. This general factor may be a component of the distinction between classicism and romanticism. In nonrepresentative designs, which characterize the majority of the art works in this sample, romanticism might be described in terms of complexity of design and specific variables such as asymmetry, curved lines and crowded figures. Most of these 11 variables have been included in a list of variables described by C. Sachs (3) as related to the classic-romantic distinction.

The oral and dependence systems of behavior, shown in Table 2 to have the highest relationships to complexity of design in art, are mainly concerned with adequacy of food, affection, and protection. In the majority of cultures with complex art style, to draw an implication from this finding, the typical individual learns self-reliant behavior to a high degree and is punished or frustrated for overt expression of dependence. On the other hand, the systems of behavior with lower relationships to art style (anal, sexual, and aggression) are concerned with prohibition or restriction of pleasurable actions, and imply pressure toward obedient and compliant behavior. The correlation of complex art style with severe socialization thus apparently applies primarily to severe socialization pressures toward independent behavior rather than toward obedient behavior.

The correlations reported in this paper are subject to various alternative interpretations. Such ambiguity can best be reduced by determining with more certainty the meanings of the measures of socialization and art style, as by finding the relationships of these measures to independent measures of personality, aesthetic expression, and other aspects of cultural practices.

SUMMARY

Works of pictorial art of 30 nonliterate cultures were rated on a number of variables of art style. Eleven of these art variables were found to be related to each other on a dimension of complexity of style. The art variables were correlated with a combined measure of severity of socialization from ratings presented by Whiting and Child (4). For each of the 11 art variables the cultures above the median in complexity tended to be above average in severity of socialization; for two of the art variables the correlation was statistically significant. The variable entitled complexity of design, which was more closely related than any other art variable to the overall measure of complexity, showed a higher correlation than any of the others with the measure of severity of socialization. Cul-

tures above the median in complexity of design tended to be above average in each of the 10 separate measures of socialization severity.

A possible interpretation of the results is that a personality characteristic is related to severity of socialization and complexity of art style in the individual, and that the cultural custom which perpetuates a pattern of socialization and art style is an adjustment to the personalities of the individuals in the culture.

REFERENCES

1. DAVIDOFF, M. D., & GOHEEN, H. W. A table for the rapid determination of the tetrachoric correlation coefficient. *Psychometrica*, 1953, **18**, 115–121.
2. FINNEY, D. J. The Fisher-Yates test of significance in 2 × 2 contingency tables. *Biometrika*, 1948, **35**, 145–156.
3. SACHS, C. *The commonwealth of art*. New York: W. W. Norton, 1946.
4. WHITING, J. W. M., & CHILD, I. L. *Child Training and personality*. New Haven: Yale Univer. Press, 1953.

Received July 10, 1956.

THE GENERALITY OF ACCURACY AND RESPONSE SETS IN INTERPERSONAL PERCEPTION[1]

WAYMAN J. CROW AND KENNETH R. HAMMOND[2]

University of Colorado

PSYCHOLOGISTS have often expressed the belief that a successful adjustment to a social environment depends upon an accurate perception of other people in the environment. This perceptual process has been called "empathy," "social perceptiveness," "social sensitivity," and "insight into others," but will here be called *interpersonal perceptiveness*, according to current usage (cf. 6).

Previous students of this problem have followed formally similar procedures. Judges (subjects), after observation of some other individual or group (objects), are asked to estimate some characteristic of the objects[3] for which some operational measure is available. This measure is then used as a criterion to determine the accuracy of the *S*s. Although a wide variety of procedures have been used within this framework, three main types may be distinguished: (*a*) estimation of responses to personality inventories, (*b*) estimation of responses to questionnaires, and (*c*) estimation of ratings within small groups.

Studies using these techniques are indeed apparently similar--but do they produce similar results; i.e., are results generalizable? Taft (10) has pointed out the paucity of studies comparing these techniques, and Gage and Cronbach (6) note that many writers assume that a general ability or trait of interpersonal perceptiveness exists. If such a general trait does in fact exist, then results obtained using different techniques are comparable. If no such general trait is established, then greater caution than some writers have exercised will be necessary when results are compared.

Moreover, any study of generality must take into account the fact that as knowledge has accumulated about the judgments that *S*s make of others on the basis of brief observation, it has become apparent that what the *S* "brings to the situation" determines to a marked degree his response to the task. In attempts to sharpen concepts and to eliminate artifactual components of accuracy scores, attention has been called to the *S*'s tendencies to "project" (8), to "assume similarity" (4), and to have "response sets" (3) and "implicit personality theories" (2). That such artifactual components invalidate conclusions regarding interpersonal perception accuracy has been pointed out in Campbell (1), Cronbach (2), and Crow (3). Such subject-determined (as distinct from situational- or object-determined) characteristics, in addition to their effect upon accuracy, may have psychological interest in their own right and may, in fact, be of more significance than accuracy in certain instances. For example, Fiedler has demonstrated the significance of "assumed similarity" for the prediction of team effectiveness (5).

This paper is a report of two studies. The purpose of the first study was to investigate the generality of interpersonal perceptiveness over *different procedures* when the influence of response sets upon the accuracy scores was controlled statistically. A variety of interpersonal perception measures was administered to a group of Senior medical students. Their scores on these different measures were intercorrelated in order to determine the extent to which generality over measures exists.

The purpose of the second study (involving different *S*s) was to investigate the generality of the accuracy of interpersonal perception *over time* when the effects of response sets have been eliminated, and, in addition, to investigate the generality-over-time of certain response-set characteristics themselves. The same measures of interpersonal perceptiveness were administered three times, at six-month intervals, to a single group of *S*s.

[1] This investigation was carried out as part of a larger study of medical education supported by the Commonwealth Fund of New York.

[2] The authors are indebted to Fred Kern, Jr., M.D., for his assistance during many stages of the study.

[3] To avoid confusion those persons who make judgments about others are called judges or subjects; those persons about whom judgments are made are called objects.

METHOD

The Scoring of Measures of Interpersonal Perception

Conventionally, interpersonal perception tests have been scored by finding the difference between the judge's estimate and the object's performance. These differences are then summed for all the estimations that the judge made about the object. Critical analysis of this procedure has revealed that such scores contain components that may have little to do with accuracy. Gage, Leavitt, and Stone (7) suggest the use of "intermediary scoring keys" and Cronbach (2) has proposed an analysis of variance method of scoring that eliminates the effects of these components upon the accuracy score and yields a highly refined measure of "differential accuracy." "Differential Accuracy reflects ability to predict differences between *O*'s on any item" (2, p. 179). The author (3) has proposed a Random-Comparison Method of determining differential accuracy which differs from Cronbach's method in that it is distribution-free, or "nonparametric." Cronbach's method has the advantages of greater precision and scope whenever the assumptions of normality of distribution of the judge's estimations and the object's scores are warranted. The Random-Comparison Method has the advantage of usefulness when assumptions of normality are not justified. Although the latter method may be less time consuming to compute, this advantage is lost whenever it is desirable to obtain separate measures of the various response set components. This objective is accomplished easily with Cronbach's mathematical model, but requires separate computations with the Random-Comparison Method.

In the present study the Random-Comparison Method was used to eliminate artifactual components of the scores. Whenever the judges make estimations about several objects, a distribution of comparisons is generated by comparing every estimation with every object's score, ignoring the signs. Thus, if eight judges each make an estimation for each of ten objects, the comparison procedure would yield 800 difference scores. The difference score obtained by comparison of the judge's estimation with the score of the object for whom the estimation was intended can be compared with the median of the random comparison distribution. If this difference score is more accurate (lower) than the median of the random comparisons, then the judge has obtained differential accuracy with that object. The final score is the percentage of objects upon which the judge was more accurate than the median of the random comparison distribution. This procedure can be further refined by using only the estimations of the individual judge to generate a random comparison distribution instead of pooling the estimations of all judges (which was done with most measures in this study[4]).

Measures: Movie Case Test

Sound motion pictures were taken of a physician interviewing a patient. The patients were selected at random from the medical wards of the Colorado General and Denver General Hospitals. Only those patients who were well enough to be moved to the room in which the filming took place, who had no language difficulty, and who were neither senile nor psychotic were selected. After the six-minute film was taken, the patients were administered the vocabulary section of the Wechsler-Bellevue Test, the Minnesota Multiphasic Inventory (card sort form), a reticence test, and a self-rating scale for a number of variables which were rough paraphrases of the MMPI Scales.

The task presented to the *Ss* was to estimate the patient's responses to the above measures after viewing the six-minute sound film. Thirty-six films were available, from which ten were selected at random. Each of the tests was scored by the Random-Comparison Method using the pooled estimations of all *Ss* to generate the difference scores as outlined below.

Reticence test. This measure concerned the reticence of the patients. The patients were asked to sort 48 cards containing statements about topics such as "your friends," "your cooperation with your doctor," "your attitude toward homosexuals," and so on. The patient was asked to sort these cards into two categories—those topics he was willing to discuss with the doctor in a subsequent interview in one category, those he was unwilling to discuss in another category. The entire list of topics and the actual instructions given to the patients were presented to the *Ss* in the test form. Their task was to estimate the number of topics that the patient was unwilling to discuss, i.e., the amount of the patient's reticence.

Vocabulary tests. The purpose of these measures was to discover how accurately the *Ss* could judge the vocabulary level of a patient from the film. The patients were administered the Vocabulary Section of the Wechsler-Bellevue Test. The list of words as well as the instructions given to the patients were presented to the students. They were asked to estimate the first word that the patient missed (Test A) and the three words in a row the patient missed (Test B).

Personality scales. The patients were administered the MMPI (9) and a series of 7-point self-rating scales that were paraphrases of the MMPI scales. The students were given two tasks: to estimate the point on each of the seven scales where the patient rated himself (Test A), and to estimate the point on the scales where the patient "really is" (Test B). The actual self-ratings served as criteria for the first task and the MMPI profile as a criterion for the second. The following 7-point scales were used by the patients for the self-ratings; the corresponding MMPI scales are in parentheses:

Scale 1 (Lie Score). How much are you concerned with what the doctor thinks of you as a person: extremely concerned, very concerned, more than most people, about the same as most people, less than most people, unconcerned, don't care at all?

Scale 2 (K Score). Do you feel that you are: extremely critical of yourself, overly critical of yourself, more critical of yourself than most people, just as critical as most people, less critical than most people, not very critical of yourself, never critical of yourself?

Scale 3 (Hypochondriasis). How do you feel about your health and your past illnesses? Have you had:

[4] For a more detailed description of this procedure and a discussion of the types of test situations for which it is applicable, the reader is referred to (3, pp. 38–43).

more trouble than anyone you know, far more than most people, more than most people, about as much as most people, less than most people, far less than most people, less than anyone you know?

Scale 4 (Paranoia). When you have had dealings with people, how have you found them to be? Was there: no one you could really trust, a few you could trust, some you could trust, you could trust about half the people, more than half you could trust, almost all could be trusted, no one you couldn't trust?

Scale 5 (Psychasthenia). Some people are depressed, worry too much, have a lack of confidence, and are unable to concentrate. Would you say that you are usually: not at all like this, very little like this, some of it fits—most of it doesn't, some of it fits—some of it doesn't, most of it fits you, fits you well, fits you perfectly?

Scale 6 (Hypomania). How active a person have you been? Have you been engaged in: very many different activities, many activities, quite a few, some, not many, a few, very few different activities?

Scale 7 (Social Contact). Do you have your best times when you are: with a large number of strangers, with a large number of people—many of them strangers, with a large number of friends and strangers, with some friends and some strangers, with a small group of friends, with a few close friends, alone?

Measures: Sociometrics Test

The purpose of the Sociometrics Test was to measure the *Ss*' ability to estimate ratings in small groups. *Ss* were assigned at random to groups of ten members each, and were asked to place everyone in the group, including themselves, in rank order on the traits of leadership, cooperation, and likeability. This was called Task A. They were then asked to estimate where each individual would be ranked by the group as a whole for the same three traits (Task B). They were next asked to estimate how each of the nine others in the group had ranked them on these traits (Task C). And, finally, they were asked to estimate how each of the others in the group had ranked himself on the three traits (Task D). As criteria for the *Ss*' accuracy in each of the above tasks, the actual rank orders (Task A) were available. That is, Task A provided the group rank order, the way each person in the group ranked each *S*, and the way each person ranked himself.

Measures: Estimates of Group Opinion Test

All *Ss* answered a questionnaire concerning their attitudes toward medical education, and were then asked to estimate the percentage of the class who answered each question affirmatively. In all, 26 questions were used. Each student's accuracy was determined by correlating his estimated percentages with the actual percentage of the class who answered the corresponding questions affirmatively.

STUDY I

THE GENERALITY OF ACCURACY IN INTERPERSONAL PERCEPTION

The generality of accuracy of interpersonal perception over different procedures was studied by administering the measures described above to a group of 65 Senior medical students. The Random-Comparison Method of scoring was used for the *Ss*' estimations of how each of the others in the group had ranked them (Task C) and for their estimations of how each of the others in the group had ranked himself (Task D). The estimations of all the *Ss* were not pooled as in the Movie Case test measures, but a separate random comparison distribution was computed for each *S*. This procedure was necessary because each *S* made estimations about the nine others in his group, who were therefore differently composed for each *S*. It was not necessary to employ the Random-Comparison Method with the estimation of the group rank order (Task B). Because the *Ss* were required to estimate a rank order, the estimations were necessarily spread evenly over the scale, and because the objects were assigned ranked scores, these scores were also evenly distributed across the scale. This procedure eliminates the effects of response set upon the *Ss*' accuracy scores. A difference score was computed between the estimated and actual group rank orders.

Results and Discussion

Table 1 presents the means and standard deviations of the accuracy scores. Apparently a sufficient range of ability was present to permit significant correlations among the ac-

TABLE 1

MEANS AND STANDARD DEVIATIONS OF ACCURACY SCORES*

Measure	Mean	*SD*
Reticence Test	44.77	12.2
Vocabulary		
Test A	68.92	19.8
Test B	72.62	14.5
Personality Scales		
Test A	65.85	6.1
Test B	62.55	9.1
Sociometrics		
Task B Lead.	19.63	6.8
Coop.	22.54	6.7
Like.	19.95	7.2
Task C Lead.	63.97	13.7
Coop.	65.89	12.9
Like.	67.91	13.7
Task D Lead.	67.43	12.1
Coop.	58.12	13.6
Like.	61.20	14.9
Group Opinion (Z-Score)	54.86	22.0

* $N = 65$.

TABLE 2
CORRELATION COEFFICIENTS AMONG MEASURES OF SOCIAL PERCEPTION ACCURACY*

Measure	Vocabulary		Personality Scales		Sociometrics									Group Opinion Z-Score
					Task B			Task C			Task D			
	A	B	A	B	Lead	Coop	Like	Lead	Coop	Like	Lead	Coop	Like	
Reticence Test	.19	.09	−.05	.02	−.06	−.09	−.33†	.07	−.21	.01	.13	−.08	.02	.09
Vocabulary														
Test A		.19	−.13	−.00	−.15	−.11	−.18	.10	−.06	−.06	.25†	−.07	.25†	−.17
Test B			−.07	−.07	.06	−.05	.06	.10	.05	−.04	.25†	−.13	.06	−.04
Personality Scales														
Test A				.18	−.03	−.23†	−.04	.04	.11	.41†	.08	−.05	−.04	.06
Test B					.06	−.09	.07	.14	−.08	.25†	.00	−.00	.09	−.05
Sociometrics														
Task B Lead.						.15	.13	.01	−.08	−.19	−.10	−.11	.02	−.21
Coop.							.39†	−.02	.00	−.12	−.06	.01	−.12	−.15
Like.								−.21	−.16	.08	.05	−.04	−.19	−.19
Task C Lead.									.08	.12	.19	.03	.01	−.03
Coop.										.19	−.18	−.15	.04	.04
Like.											.12	.13	−.05	.14
Task D Lead.												.08	.02	−.07
Coop.													.08	.11
Like.														−.09

* $N = 65$.
† Significant beyond the .05 level.

curacy scores, although the parameters of ability in interpersonal perception are as yet unknown.

Table 2 presents the product-moment correlation coefficients among the 15 measures of social perception accuracy. Only eight of these 105 correlations are significant beyond the .05 level of confidence, and two of these are negative relationships. These results are interpreted as failing to support the hypothesis of a general ability of interpersonal perceptiveness. Apparently interpersonal perceptiveness is more specific than previously assumed. Even though generality may be demonstrated with other measures and/or other *S*s, these negative results make the *assumption* of generality untenable. Consequently, comparison of the results of interpersonal perception studies based on different measuring techniques can be justified only when their comparability has been empirically demonstrated.

STUDY II

THE GENERALITY-OVER-TIME OF ACCURACY AND RESPONSE SETS

The generality-over-time of accuracy and response sets was investigated by administering the Personality Scales test (see above) three times, at six-month intervals, to the same group of *S*s.

The *S*s (not the same as in Study I) were 72 Senior medical students at the University of Colorado Medical School, who were tested at the beginning of their Senior year (Pretest), the middle of the year (Posttest-I), and at the end of the year (Posttest-II). The same measures were used, but the *S*s made estimations for three different sets of ten patients.

Differential accuracy scores that were free of the influence of response sets were computed by use of the Random-Comparison Method as in Study I. In addition, three response-set characteristics were investigated in the present study:

1. The *implicit stereotype score*, defined as the mean of an *S*'s estimations for all the objects on the same scale. For example, *S*s were asked to estimate the self-ratings of ten patients on a seven-point scale "concern for what the doctor thinks of you as a person." The scale ran from "extremely concerned" to "don't care at all." An *S* may think that people *in general* are little concerned, or he may think that they are greatly concerned; this "stereotype" will determine to a certain degree the position on the scale around which

the perceived individual differences among the patients will cluster. We have used the *mean* of the estimations as an indicator of this stereotype.

2. The *adherence to stereotype score*, defined as the inverse of the average variance of an *S*'s estimations for the scales. An *S* may perceive wide differences among people in the characteristic under consideration, or, he may perceive people as being very much alike. He may, in other words, cluster his estimations closely around his stereotype (mean estimation) or he may vary them widely. The variance of the estimations for each scale was computed and averaged for the seven scales as a measure of perceived individual differences.

3. The *assumed veridicality score*, defined as the inverse of the average difference between the *S*'s estimation for the patient's self-rating and his estimation as to where the patient "really was." An *S* could assume a large measure of agreement between a patient's self-rating and "reality," or he could assume little agreement.

The *S*s' stereotype, adherence to stereotype, and assumed veridicality scores were computed for each of the three administrations. The scores on each scale were combined to furnish composite scores for adherence to stereotype and assumed veridicality. The implicit stereotype scores were not combined but were treated separately since it would be expected that the stereotype would be unique for each scale. The scores on each administration were correlated with the scores on the other two administrations to determine the relationship between performance at one time and performance at the end of two successive six-month intervals.

Results and Discussion

Table 3 presents the product-moment correlation coefficients among the stereotype scores of the students on the Pretest, Posttest-I, and Posttest-II administrations. Of these 48 correlations only two *failed* to reach significance at the .01 level of confidence. Thus, the stereotypes that students hold regarding the patients on these scales were remarkably persistent over a six-month and one-year period.

Table 4 presents the product-moment correlation coefficients among the differential accuracy, adherence to stereotype, and assumed veridicality scores for three administrations to the same *S*s at six-month intervals.

TABLE 3

PRODUCT-MOMENT CORRELATION COEFFICIENTS AMONG THE STEREOTYPE SCORES FOR THREE ADMINISTRATIONS*

Measure	Pretest vs. Posttest I	Pretest vs. Posttest II	Posttest I vs. Posttest II
Task A			
Scale 1	.51	.54	.56
Scale 2	.51	.30	.36
Scale 3	.51	.45	.52
Scale 4	.58	.47	.41
Scale 5	.39	.54	.55
Scale 6	.59	.46	.66
Scale 7	.56	.47	.65
Scale 8	.52	.30	.47
Task B			
Scale 1	.44	.51	.63
Scale 2	.54	.68	.45
Scale 3	.45	.46	.54
Scale 4	.61	.36	.32
Scale 5	.57	.63	.65
Scale 6	.54	.54	.51
Scale 7	.67	.41	.43
Scale 8	.54	.44	.56

* $N = 72$ and $r = .31$ at the .01 level of confidence.

TABLE 4

PRODUCT-MOMENT CORRELATION COEFFICIENTS* AMONG ACCURACY, ADHERENCE TO STEREOTYPE, AND ASSUMED VERIDICALITY SCORES FOR THREE ADMINISTRATIONS

Measure	Pre-Post I	Pre-Post II	Post I-Post II
Accuracy			
Task A	.19	.10	.24
Task B	.23	.28	.38
Adherence to Stereotype			
Task A	.47	.55	.66
Task B	.53	.50	.74
Assumed Veridicality	.64	.63	.76

* $N = 72$ and $r = .31$ at the .01 level of confidence.

Only one of the six correlations for the differential accuracy scores is significant at the .01 level of confidence and three are significant at the .05 level. While there seems to be a positive relationship among accuracy scores obtained at six-month and one-year intervals, the relationship is not a strong one as compared to the relationships indicated by the correlations for adherence to stereotype and assumed veridicality scores. All of these latter correlations are significant beyond the .01 level of confidence. The *largest* correlation for

accuracy was .38, while the *smallest* for the adherence to stereotype and assumed veridicality correlations was .47.

It should be remembered that while the format and content of the estimations which constituted the task for the *S*s were identical, three different sets of films of ten patients each were administered. Furthermore, a full year of medical school activity intervened between performance on the Pretest and performance on Posttest-II, and six months intervened between the other administrations. The obtained correlations for stereotype, adherence to stereotype, and assumed veridicality are considered to be very high in view of these circumstances.

The relationship between performance at one time and performance six months and one year later that was found for the three response sets indicates that these characteristics are enduring and stable aspects of the *S*s' approach to their task. The lower correlations that were obtained for the differential accuracy scores indicate that they are not as stable as response sets, in this situation at least. If it is true in general that, in the interpersonal perception task, response sets are more stable aspects of the *S*s' performance than differential accuracy, then consistency in their performance may be due to consistency in response sets rather than to consistency in interpersonal perceptual ability. Since investigators of interpersonal perception have rarely eliminated response-set components from their accuracy scores, they are subject to precisely the above criticism. The reliabilities reported for accuracy of interpersonal perception should be seriously questioned because the consistency of the scores may be due to consistency in response set rather than to consistency in ability. Consequently, the authors maintain that differential accuracy of interpersonal perception as a reliable characteristic of *S*s' performance *is yet to be established.*

Response sets deserve separate investigation aside from their influence upon accuracy scores. The *S*s in this study had consistent modes of response to the judgmental task—they were consistent over time in using implicit stereotypes, consistent in the extent to which they adhered to these stereotypes, and consistent in the degree to which they assumed veridicality between the patient's self-rating and "reality."

At present an investigation is being made of the personality correlates of such consistent response tendencies, and an attempt is being made to discover whether they have relevance for training in interpersonal relations.

SUMMARY

Two studies of the generality of interpersonal perceptiveness are reported. The first was an investigation of accuracy as a general ability over various procedures and the second was an investigation of the stability of accuracy scores and certain response-set scores over time. Differential accuracy was defined as the ability of an *S* to predict differences among other people (objects) on any item. In order to eliminate the influence of response sets upon the differential accuracy scores the Random-Comparison Method was used.

In the first study fifteen measures of interpersonal perception accuracy were administered to 65 Senior medical students. These measures were obtained from three procedures: (*a*) estimations of responses to personality inventories; (*b*) estimates of group opinion; and (*c*) estimation of ratings within small groups. These scores were intercorrelated to determine the generality of interpersonal perceptiveness, i.e., to determine whether or not subjects who were good (or poor) on one measure were good (or poor) on other measures. Only eight of the 105 correlations were significant at the .05 level of confidence and two of these were negative. The results were interpreted as failing to support the hypothesis of a general ability of interpersonal perceptiveness.

In the second study (involving a different group of *S*s) the same measures of interpersonal perceptiveness were administered to 72 Senior medical students three times at six-month intervals. Thirty-six patients were selected at random (with certain restrictions) from the medical wards of two hospitals and were interviewed by a physician. Thirty-six, six-minute sound movies were taken of these interviews and ten of them were presented to the *S*s at each of three test administrations. *S*s were required to estimate the self-ratings of the patients and the point on the scale where the patient "really was" for seven per-

sonality scales. The criteria for accuracy were the actual self-ratings of the patients and the patients' performance on the MMPI.

The Random-Comparison Method was used to compute differential accuracy scores free of artifactual components. Three "response-set" scores were determined: (*a*) implicit stereotype score—the mean of an *S*'s estimations, (*b*) adherence to stereotype score—the inverse of the variance in his estimations, and (*c*) assumed veridicality score—the inverse of the difference between estimations for self-rating and for the patient's "real" score, summed for all seven scales and ten patients.

The scores on each administration were intercorrelated with similar scores on the other two administrations. The results indicated (*a*) significant but low relationships among accuracy scores over time, and (*b*) significant and high relationships among the response-set scores over time. These findings were interpreted to mean that response-sets were more stable elements of the *S*'s performance than differential accuracy and that previously reported reliabilities are spuriously high whenever the influence of response sets upon the accuracy scores has not been eliminated.

The conclusions from the two studies were as follows:

1. Failure to find support for the hypothesis of a general ability of interpersonal perceptiveness in this study makes the *assumption* of generality untenable. Consequently, comparison of the results of interpersonal perception studies based on different measuring techniques can be justified only when the comparability of such results has been empirically demonstrated.

2. The finding that response sets were more consistent over time than differential accuracy indicates that the reliabilities reported for accuracy of interpersonal perception should be seriously questioned because the consistency of the scores may be due to consistency in response set rather than to consistency in ability. Since investigators have rarely eliminated response-set components from their accuracy scores, it is maintained that differential accuracy of interpersonal perception as a reliable characteristic of performance is yet to be established.

3. The high consistency of the response sets investigated in this study indicates that they deserve separate investigation aside from their influence upon accuracy scores.

REFERENCES

1. CAMPBELL, D. T. An error in some demonstrations of the superior social perceptiveness of leaders. *J. abnorm. soc. Psychol.*, 1955, **51**, 694–695.
2. CRONBACH, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychol. Bull.*, **52**, 177–193.
3. CROW, W. J. A methodological study of social perceptiveness. Unpublished doctor's dissertation, Univer. of Colorado, 1954.
4. FIEDLER, F. E. A method of objective quantification of certain counter-transference attitudes. *J. clin. Psychol.*, 1951, **7**, 101–107.
5. FIEDLER, F. E. Assumed similarity measures as predictors of team effectiveness. *J. abnorm. soc. Psychol.*, 1954, **49**, 381–388.
6. GAGE, N. L., & CRONBACH, L. J. Conceptual and methodological problems in interpersonal perception. *Psychol. Rev.*, 1955, **62**, 411–422.
7. GAGE, N. L., LEAVITT, G. S., & STONE, G. C. The intermediary key in the analysis of interpersonal perception. *Psychol. Bull.*, 1956, **53**, 258–266.
8. HASTORF, A. H., & BENDER, I. E. A caution respecting the measurement of empathic ability. *J. abnorm. soc. Psychol.*, 1952, **47**, 574–576.
9. HATHAWAY, S. R., & MEEHL, P. E. *An atlas for the clinical use of the MMPI.* Minneapolis: Univer. of Minnesota Press, 1951.
10. TAFT, R. The ability to judge people. *Psychol. Bull.*, Vol. **52**, 1–23.

Received July 10, 1956.

PATTERNS OF ATTITUDES TOWARD CHILD REARING AND THE FAMILY

EARL S. SCHAEFER AND RICHARD Q. BELL

National Institute of Mental Health, National Institutes of Health, Public Health Service, U. S. Department of Health, Education, and Welfare

RESEARCH interest in the influence of the parent-child relationship upon personality development of the child has been stimulated by studies (4, 5, 8, 9, 11, 12, 17) which have reported that objectively measured attitudes of parents toward child rearing are related to emotional and social adjustment of children. Since it has been assumed that the behavior of parents with their children is a manifestation of personality functioning, Block's study (3), which demonstrated relationships between attitudes toward child rearing and personality characteristics of fathers, lends additional support to these findings. Clinical studies of mothers of maladjusted children, which have been reviewed by Symonds (18) and Radke (11), have also revealed syndromes of maternal behavior and attitudes which are related to differential personality structure of children. In the attitude studies cited, a number of parental concepts have been explored, but studies of a comprehensive sample of attitudes toward child rearing have not as yet been reported.

This study extends a series of investigations which have developed an objective inventory of attitudes toward child rearing (2, 14, 15) and a set of rating scales of maternal personality and of mother-child interaction (16). These research instruments are based upon an analysis of the domain of mother-child interaction in terms of concepts which are hypothetically related to the behavior of a mother with her child as it may affect the personality development of the child. Previous studies have demonstrated the feasibility of measuring the attitudes and behavior of women who are actually involved in relationships with their children. The present study was planned to determine whether it is possible to differentiate meaningful patterns of maternal attitudes in young unmarried women. Success in isolating theoretically significant patterns of maternal attitudes in this population would provide a basis for an attempt to predict mother-child interaction from attitude data gathered prior to the birth of a child.

METHOD

The Parental Attitude Research Instrument (2, 14), a questionnaire which had been developed to obtain multivariate measurement of attitudes toward child rearing and the family, was used to explore this domain. The 24 Likert-type attitude scales are listed in Table 1 with one sample item from each eight-item scale.

The subjects were 100 unmarried student nurses in their freshman and sophomore years. The homogeneity of the group in socioeconomic background, occupational goal, educational level, age, and intelligence permitted the control of a number of cultural factors to which variability in attitude might be attributed. Scores on the Otis Higher Intelligence Test which were available on a chance sub sample of 65 members revealed a median IQ of 111 with a range from 95 to 127. The inventory was given as an anonymous objective group test in regularly scheduled classes of psychology. *Ss* were asked to indicate whether they strongly agreed, mildly agreed, mildly disagreed, or strongly disagreed with each item. Scores for each scale were computed by summing the responses which had been converted into score weights of 1 through 4. The internal consistency reliability coefficients, computed with Kuder-Richardson Formula 20, are given in Table 2. A centroid factor analysis was performed on the product-moment inter-scale correlations, and the factors rotated obliquely to obtain simple structure (19).[1]

RESULTS AND DISCUSSION

Despite the relatively low reliabilities of the scales, probably due to the homogeneity of the sample as well as the relatively short scales, the reliabilities were considered adequate for an investigation of the factorial structure of this domain. The rotated factor loadings on the five relatively independent factors which were isolated are given on Table 2. Scales which have their highest loading on a given factor are grouped under a tentative description or name of the factor which is given in italics. Since several scales have loadings above .30 on two factors, a

[1] The correlation matrix, the centroid factor matrix and the transformation matrix have been deposited with the American Documentation Institute. Order Document No. 5137 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 1

NAMES OF SCALES WITH A SAMPLE ITEM FROM EACH

1. Encouraging Verbalization:
   Children should be allowed to disagree with their parents if they feel their own ideas are better.

2. Infantilization:
   A well-trained child always keeps the same close bond with its mother.

3. Seclusion of the Mother:
   Too many women forget that a mother's place is in the home.

4. Breaking the Will:
   Children need some of the natural meanness taken out of them.

5. Harsh Punishment:
   If small children refuse to obey, parents should whip them for it.

6. Fostering Dependency:
   A mother should do her best to avoid any disappointment for her child.

7. Martyrdom:
   Children should realize how much parents have to give up for them.

8. Marital Conflict:
   Sometimes it's necessary for a wife to tell off her husband in order to get her rights.

9. Strictness:
   Strict discipline develops a fine strong character.

10. Irritability:
    Children will get on any woman's nerves if she has to be with them all day.

11. Excluding Outside Influences:
    Children should never learn things outside the home which make them doubt their parents' ideas.

12. Deification:
    The child should be taught to revere his parents above all other grown-ups.

13. Suppression of Aggression:
    There is no good excuse for a child hitting another child.

14. Rejection of the Homemaking Role:
    Having to be with the children all the time gives a woman the feeling her wings have been clipped.

15. Autonomy of the Child:
    Children should be given a chance to try out as many things on their own as possible.

16. Avoidance of Communication:
    If you let children talk about their trouble they end up complaining even more.

17. Ignoring the Baby:
    Handling a baby too much makes him spoiled and demanding.

18. Suppression of Sexuality:
    It is very important that young boys and girls not be allowed to see each other completely undressed.

19. Intrusiveness:
    An alert parent should try to learn all her child's thoughts.

20. Ascendance of the Mother:
    A married woman knows that she will have to take the lead in family affairs.

21. Comradeship and Sharing:
    If more parents would have fun with their children, the children would be more apt to take their advice.

22. Demand for Activity:
    The sooner a child learns that a wasted minute is lost forever the better off he will be.

23. Abdication of the Parental Role:
    Some children are so naturally headstrong that a parent can't really do much about them.

24. Acceleration:
    A mother should make an effort to get her child toilet trained at the earliest possible time.

TABLE 2
ROTATED FACTOR MATRIX OF PARENTAL ATTITUDES

Factors with Scales Defining Each Factor	Reliability*	Factor Loadings				
		I	II	III	IV	V
I. Suppression and Interpersonal Distance						
21 Comradeship and Sharing	62	*−61*	06	20	−03	06
15 Autonomy of the Child	61	*−57*	24	20	−24	13
18 Suppression of Sexuality	58	*42*	01	04	*40*	−05
16 Avoidance of Communication	59	*30*	09	19	−05	22
1 Encouraging Verbalization	61	*−54*	*51*	01	03	−16
II. Hostile Rejection of the Homemaking Role						
14 Rejection of the Homemaking Role	72	01	*64*	17	15	−03
23 Abdication of the Parental Role	64	00	*61*	00	23	00
10 Irritability	73	07	*60*	04	−19	30
8 Marital Conflict	64	−14	*47*	04	−04	17
17 Ignoring the Baby	77	27	*38*	22	−06	03
III. Excessive Demand for Striving						
9 Strictness	74	−07	07	*58*	−19	*35*
24 Acceleration of Development	64	00	05	*51*	10	00
22 Demand for Activity	66	*−34*	26	*49*	11	14
IV. Over-possessiveness						
7 Martyrdom	65	−09	−03	*36*	*59*	−07
13 Suppression of Aggression	54	00	08	−01	*57*	13
6 Fostering Dependency	60	19	06	−14	*53*	13
19 Intrusiveness	84	03	−05	12	*52*	01
2 Infantilization	57	15	18	02	*49*	16
V. Harsh Punitive Control						
11 Excluding Outside Influences	67	06	−04	*41*	−02	*55*
4 Breaking the Will	63	04	16	26	−01	*53*
20 Ascendance of the Mother	74	*44*	08	00	00	*52*
12 Deification of the Parent	79	−03	−13	*41*	16	*45*
3 Seclusiveness of the Mother	61	23	−16	−05	17	*42*
5 Harsh Punishment	76	13	17	15	−27	*39*

* Internal Consistency Reliability Coefficient—Kuder-Richardson Formula 20.

scale may be mentioned twice in the description of factors below.

The factor of Suppression and Interpersonal Distance is defined by both negative and positive loadings. The attitudes of Comradeship and Sharing, Autonomy of the Child, and Encouragement of Verbalization seem to reflect approval of a close relationship with a child which still permits him freedom to develop as a differentiated individual and also to express his views about family affairs and parent behavior toward him. The other pole of this factor reflects an attitude that the mother should dominate the family, that the sexual interests of the child should be tabooed, that children should not be allowed to talk about their fears and hostilities with their parents, and that a baby needs only physical care without social play and tenderness from the mother. This factor seems to be related to the warmth and closeness of the relationship and the degree of tolerance for expression of individuality of the child.

The content of the factor, Hostile Rejection of the Homemaking Role, reveals dissatisfaction with the role of housewife and mother and a high level of hostility which is openly expressed. The scale, Rejection of the Homemaking Role, consists of items which state the unhappiness of a woman at her loss of freedom and her dissatisfaction with the duties of caring for a home and children. An attitude that one can do very little in rearing children, with an implied denial of the responsibility of a parent, is revealed in Abdication of the Parental Role. Irritability items imply impatience, intolerance, and a low threshold for the arousal of hostile reactions toward children. In the scale Encouraging Verbalization the *S* reveals her approval of a child's open expression of his disagreements with and hostility toward his parents. Marital Conflict reflects

an expectation of a great deal of quarreling and fighting between husband and wife and again reflects hostility which is openly expressed. Items on Ignoring the Baby express the opinion that a baby needs only physical care and that much of the attention given to him is unnecessary. This factor may be related to the pattern of maternal rejection which has been described by Newell (10), Symonds (18) and others.

The factor, Excessive Demand for Striving, is defined by scales which reveal a demand for conformity and achievement by the child to meet the parents' needs. Strictness demands conformity to rigid rules; Acceleration of Development states that a child should reach developmental goals at an accelerated pace; and Demand for Activity items reflect a belief that a child should be kept busy and should achieve as much as possible. Exclusion of Outside Influences, Deification of the Parent, Martyrdom, and Breaking the Will may be related to methods of enforcing these demands for conformity and achievement. This factor to some degree resembles Lorr and Jenkins' (7) factor, "organization and effectiveness of control in the home," derived from a second order factor analysis of the Fels Parent Behavior Rating Scales.

The factor of Over-Possessiveness seems to have a central theme of control of the child through keeping the child indebted to the mother, dependent, and immature. The highest loading, on Martyrdom, reveals a perception of a mother as a sacrificing, suffering woman with an implied obligation of the child to feel guilty and indebted to her. Suppression of Aggression indicates a wish for a quiet child who would be more easily controlled than an active, aggressive one. Infantilization refers to a preference for an immature child who is closely tied to his mother's apron strings, and Fostering Dependency reflects a perception of the child as a weak person who must be protected by the mother. Suppression of Sexuality indicates a taboo on sexual curiosity as well as behavior. A belief that the mother should know everything the child is thinking and doing is revealed by the Intrusiveness scale. This factor resembles to some degree the factor of maternal behavior described by Roff (13) as "concern for the child," Lorr and Jenkins' (7) factor of "dependency encouraging," and Levy's (6) concept of "maternal over-protection."

The factor of Harsh Punitive Control resembles the concept of parental dominance which has been described by Symonds (18) among others. In this factor the scale, Exclusion of Outside Influences, reveals a wish for absolute control by the parents which excludes any other source of influence. Enforcement of parental authority through fear and forcing the child into submission is revealed in Breaking the Will. A need to dominate the entire family is shown by the assumption that the mother should direct all family activities. This factor also contains a scale which states that the mother should avoid an active social life, and another that approves of the use of harsh physical punishment as a method of influencing the child's behavior. The loading on Irritability reflects the extent to which the respondent feels that rearing children can be an irritating burden; and Strictness reveals an opinion that rules should be rigidly enforced. This factor of Hostile Punitive Control seems to resemble the negative pole of Roff's (13) factor of "democratic guidance" and of Lorr and Jenkins' (7) second factor of "democracy and democratic values."

Several of the factors appear to correspond to parent-child interaction patterns as described by clinicians, and to factors that have been isolated from the Fels parent behavior ratings (7, 13). The apparent similarity of these patterns arrived at from different methods of studying parent-child relationships suggests the need for studies of the degree of correlation between measures of attitudes toward child rearing and observations of actual behavior with children. The relationship between these attitudes toward child rearing and other personality variables of the respondents should also be investigated. This study, in conjunction with previous studies of objectively measured attitudes toward child rearing, suggests that prediction of mother-child interaction and of personality development of children may be possible since measurable variability in attitude and an organized structure of attitudes toward child rearing can be determined in young unmarried women.

SUMMARY

Twenty-four scales of the Parental Attitude Research Instrument were administered to 100 student nurses. Internal consistency reliability coefficients were found which permitted a centroid factor analysis of the scales. An oblique simple structure rotation revealed five relatively independent factors, identified as Suppression and Interpersonal Distance, Hostile Rejection of the Homemaking Role, Excessive Demand for Striving, Over-possessiveness, and Hostile Punitive Control. The resemblance of these attitude factors to syndromes of parental behavior and attitudes as described by clinicians in parents of emotionally disturbed children and to factors that have been isolated from the Fels Parent Behavior Scales was noted. It is suggested that this measuring instrument may permit prediction of different patterns of mother-child interaction.

REFERENCES

1. BALDWIN, A. L., KALHORN, JOAN, & BREESE, FAY. Patterns of parent behavior. *Psychol. Monogr.*, 1945, **58**, No. 3 (Whole No. 268).
2. BELL, R. Q., & SCHAEFER, E. S. Scales for measuring general orientations of parents toward children. *Amer. Psychologist*, 1955, **10**, 439. (Abstract)
3. BLOCK, J. Personality characteristics associated with fathers' attitudes toward child-rearing. *Child Develpm.*, 1955, **26**, 41–48.
4. FREEMAN, R. V., & GRAYSON, H. M. Maternal attitudes in schizophrenia. *J. abnorm. soc. Psychol.*, 1955, **50**, 45–52.
5. HARRIS, D. B., GOUGH, H. G., & MARTIN, W. E. Children's ethnic attitudes: II. Relationship to parental beliefs concerning child training. *Child Develpm.*, 1950, **21**, 169–181.
6. LEVY, D. M. *Maternal over-protection*. New York: Columbia Univer. Press, 1943.
7. LORR, M., & JENKINS, R. L. Three factors in parent behavior. *J. consult. Psychol.*, 1953, **17**, 306–308.
8. MARK, J. C. The attitudes of the mothers of male schizophrenics toward child behavior. *J. abnorm. soc. Psychol.*, 1953, **48**, 185–189.
9. MILES, KATHERINE A. Relationship between certain factors in the home background and the quality of leadership shown by children. Unpublished doctor's dissertation, Univer. of Minnesota, 1945.
10. NEWELL, A. W. The psychodynamics of maternal rejection. *Am. J. Orthopsychiat.*, 1934, **4**, 387–401.
11. RADKE, MARIAN J. Relationship of parental authority to children's behavior and attitudes. *Univ. Minn. Inst. Child Welf. Monogr.*, 1946, No. 22.
12. READ, KATHERINE H. Parents' expressed attitudes and children's behavior. *J. consult. Psychol.*, 1945, **9**, 95–100.
13. ROFF, M. A factorial study of the Fels Parent Behavior Rating Scales. *Child Develpm.*, 1949, **20**, 29–44.
14. SCHAEFER, E. S., & BELL, R. Q. Structure of attitudes toward child-rearing and the family. *Amer. Psychologist*, 1955, **10**, 319–320. (Abstract)
15. SCHAEFER, E. S., & BELL, R. Q. Parental attitude research instrument (PARI). Normative data. Unpublished manuscript. Library, National Institute of Mental Health, Bethesda, Md., 1955.
16. SCHAEFER, E. S., BELL, R. Q., & BAYLEY, NANCY. Quantification of maternal behavior and consistency of mother-child interaction. *Amer. Psychologist*, 1956, **11**, 404. (Abstract)
17. SHOBEN, E. J., JR. The assessment of parental attitudes in relation to child adjustment. *Genet. Psychol. Monogr.*, 1949, **39**, 101–148.
18. SYMONDS, P. M. *The psychology of parent-child relationships*. New York: Appleton-Century Co., 1939.
19. THURSTONE, L. L. *Multiple-factor analysis*. Chicago: Univer. of Chicago Press, 1947.

Received July 20, 1956.

AUTHORITARIAN IDEOLOGY IN NEGRO COLLEGE STUDENTS[1]

GEORGE A. STECKLER[2]

Cleveland Receiving Hospital and State Institute of Psychiatry

Most of the research stemming from *The Authoritarian Personality* (2) has concerned members of the majority group in America. Recent studies have explored ideology in the Jewish minority (1, 11, 12). The present study is an attempt to clarify the effects of minority group membership on the ideology of Negroes.

The psychological situation of the average American Negro is both ambiguous and conflict ridden. According to frustration-aggression theory (6), discrimination should arouse the Negro's aggression toward the majority group. However, the Negro may be fearful of retaliative action by the powerful majority, and he also may identify with the values of the majority, especially if he is of the middle class. He may therefore be expected to have ambivalent feelings regarding the majority group as well as his own. What are the ideological manifestations of this dilemma?

Method

Procedure and Sampling

The questionnaire method was used. The research was divided into two phases. In the first phase, preliminary forms of opinion-attitude scales were administered to 299 *Ss*, mostly members of church groups and students attending predominantly Negro colleges. The purpose of the preliminary administration was to ascertain reliability and internal consistency for the scales. On the basis of obtained data, the scales were revised and shortened.

For the study proper, *Ss* were obtained from colleges and universities attended primarily by Negroes. Cooperation was sought by letter from nine psychology instructors, all members of APA. Seven of the nine instructors responded and subsequently administered questionnaires to their classes.[3] Questionnaires were filled out anonymously. Approximately 85 per cent of the questionnaires mailed out were completed and returned. The schools from which the *Ss* were obtained were diverse in both location and character. North, South, and border states were represented. A teachers' college, an agricultural and mechanical school, and liberal arts colleges were included. Mean age of *Ss* was 20.5 years. Nearly half were Southerners; the remainder were Northerners or had lived for extended periods in both North and South. Although the group was predominantly working class in origin, family socioeconomic status was considerably higher than for the general population of Negroes.

Variables and Scales

Variables chosen for the study were: authoritarianism (F), ethnocentrism (E), politico-economic conservatism (PEC), anti-Negro ideology (AN), and anti-white ideology (AW). The California F, E, and PEC scales (2) were used to measure the first 3 of these variables. Separate scales were constructed to measure anti-Negro and anti-white ideology. Anti-Negro items were therefore omitted from the E scale. All scales were of the seven-point Likert type.

Anti-Negro and anti-white ideologies were conceived of as relatively organized and stable systems of opinions, values, and attitudes regarding Negroes and whites. They involve negative, stereotyped opinions and hostile attitudes regarding Negroes and whites. They also include moral values which justify the attitudes and permeate the opinions. In order to gain a broad understanding of anti-white and anti-Negro ideology as it exists among Negroes, Negro newspapers, magazines, and literature were perused. In formulating the items for the anti-Negro (AN) scale, it was assumed that virulent, extreme anti-Negro ideology is not common among Negroes themselves. In order to make the scale discriminating as well as nonrepugnant to the *Ss*, the items were made milder than anti-Negro ideology current among whites. However, the content of this current ideology was retained. Sixteen of the most discriminating items were selected from the preliminary form of the scale. AN items are presented in Table 1. All are phrased in the anti-Negro direction. Most of the items stress the offensiveness of Negroes and the opinion that their deviant behavior causes prejudice.

The revised anti-white (AW) scale contains 18 of the most discriminating items from the preliminary form. All but one are phrased in the anti-white direction. The items assert that whites are intrusive, untrustworthy, dishonest, and inferior in certain spheres, and advocate that they be deprived of their present power. The AW scale is presented in Table 2.

In order to shorten the length of the questionnaire and therefore economize on classroom time, two forms of the questionnaire were composed. Form A contained AN, AW, and F. Form B substituted E and PEC in place of F.[4]

[1] This article is based in part on a doctoral dissertation submitted to the Department of Psychology, Western Reserve University, February 1954. The author wishes to thank Drs. William R. Morrow and Richard W. Wallen who served as chairmen of his doctoral committee.

[2] Now at Cleveland Regional Treatment Center.

[3] The following individuals graciously consented to help: Profs. J. Otis Smith, Vera D. Hunton, Regina M. Goff, Leonora C. Lane, Paul Clifford, James F. Condell, and James T. Taylor.

[4] The difference between the means on the two scales (AN and AW) shared in common by both forms was not significant.

TABLE 1
ANTI-NEGRO SCALE ITEMS*

1. A large part of the problems facing Negroes today are caused by Negroes themselves.
2. I would rather not marry a person who has very kinky hair, wide nostrils, and thick lips.
3. The lower-class Negro is to blame for a lot of anti-Negro prejudice.
4. Whites and Negroes can get along on jobs until too many Negroes try to push themselves in.
5. One big reason why racial prejudice is still so strong is that Negroes offend people by being so sensitive about racial matters.
6. One important reason why Negroes are discriminated against in housing is that they don't keep up the property.
7. One reason why racial prejudice still exists today is the fact that many Negroes are dirty, loud, and generally offensive in their ways.
8. One trouble with Negroes is that they are even more jealous of each other's success than are whites.
9. Too many Negroes have abused the privilege of attending baseball games by being rowdy, noisy, and cheering only for the colored ballplayers.
10. Segregation and jimcrow will never end unless the average colored person becomes better educated and better mannered.
11. Colored people can hardly be expected to gain social equality until many more of them exert some effort to better themselves and live more decently.
12. With all of the drinking, cutting, and other immoral acts of some Negroes, white people are almost justified for being prejudiced.
13. Too many Negroes, when they get a little money, spend it all on whiskey, flashy cars, or expensive clothes.
14. One is almost ashamed to be a Negro when he sees so many of them who look and act like cotton pickers fresh from the fields.
15. A great many Negroes become officious, overbearing, and disagreeable when given positions of responsibility and authority.
16. Negroes would solve many of their social problems if so many of them were not irresponsible, lazy, and ignorant.

* Several items are adapted from Pitt (10).

TABLE 2
ANTI-WHITE SCALE ITEMS*

1. There is nothing lower than white trash.
2. White people may be all right, but they carry it too far when they try to butt into the Negro's affairs and go around with Negro women.
3. The whites have shown by their actions that they are naturally immoral, vicious, and untrustworthy.
4. No matter how nicely they treat a colored person, white people don't really mean it.
5. It is usually a mistake to trust a white person.
6. Any Negro who marries a white is a traitor to his people.
7. There may be a few exceptions, but white musicians and athletes are definitely inferior to Negro musicians and athletes.
8. White people are only friendly to Negroes when they want something out of them.
9. Negroes can expect no real help from white people in the fight against racial discrimination.
10. Most white people are always looking for ways to cheat and steal from the colored people.
11. The colored race has been pushed around long enough; it's about time that the whites were made to get out of the Negro communities.
12. If there is a Heaven, it is hard to imagine that there are many white people up there.
13. Although the white man now rules the world, it will be a happy day when the tables are turned and the colored people become the rulers.
14. The world might be a better place if there were fewer white people.
15. When the Bible says, "The bottom shall rise to the top,"† it gives hope that the Negro people will someday give the orders in this country instead of whites.
16. It may be wrong to damn all whites, but it's plain that whites have all the money and power, and that they look down on anyone who is colored.
17. There are many white people who are not prejudiced and who sincerely believe that Negroes are equals.‡
18. When it comes to such things as sports, dancing, music, and making love, the white man is not as talented as the Negro.

* Several items are adapted from Pitt (10).
† The Bible quotation is inaccurate.
‡ This item is positively stated: the remaining ones are negative.

RESULTS

Odd-even reliability coefficients, corrected by the Spearman-Brown prophecy formula, were computed for each scale. Table 3 presents the reliability and related statistics.

Four of the scales yielded high reliability coefficients, indicating relative freedom from chance errors of measurement and comparative stability of scores. Low reliability of PEC apparently is not solely a function of its shortness (10 items). The California study obtained much higher reliabilities for their somewhat longer version of this scale.[6] Morrow (9) reports high reliability in studies using the present PEC scale with white college students.

TABLE 3
RELIABILITY AND RELATED STATISTICS FOR THE SCALES

Property	Scale				
	E	AN	AW	F	PEC
Reliability	.75	.84	.88	.81	.44
Mean	3.58	4.55	2.75	4.66	4.39
SD	.99	1.06	1.12	.93	.75
Range	1.2–5.8	1.1–6.6	1.0–6.6	1.4–6.6	1.5–6.0
N	212	449	449	222	210
No. of Items	12	16	18	20	10

[6] By means of the Spearman-Brown prophecy formula, it was estimated that increasing the present scale to 16 items would probably increase reliability to .56, well below the *r* of .73 obtained by the California study for its 16 item PEC scale.

TABLE 4
INTERCORRELATIONS AMONG THE SCALES*

Scale	AN	AW	E	F
AN		.24	.40	.57
AW			.60	.55
E†				...

* All correlations are significant beyond the .01 level.
† The correlation between E and F was not computed because each scale was in a different form of the questionnaire.

Frequency distributions approached normality for all scales with the exception AW, which was markedly weighted with low scores and skewed in the positive direction. However, the range of scores on AW is wide and variability is greater than for any of the other scales. In spite of the bunching of low scores, the scale evidently offered sufficient opportunity for freedom of expression, agreement or disagreement.

Taking a mean score-per-item of 4.0 as the theoretical neutral point for each scale, it is apparent that the sample tends to agree with anti-Negro, authoritarian, and conservative ideologies and to disagree with ethnocentric and anti-white ideologies. The obtained means are similar in degree to those obtained in the preliminary phase of the present study. The present group's standing on AN and AW coincides with the findings of both Pitt (10) and Bayton (4), whose Negro groups are ideologically similar in spite of differences in socio-economic class. It should be noted that only one group studied by the California investigators—prison inmates—scored as high as the present sample on the F scale. However, our sample's E scale mean of 3.58 is lower than that of many of the California samples.

The correlation coefficients listed in Table 4 indicate slight to substantial relationships between scores on the several scales. This might be anticipated because each of the four scales is a measure of authoritarian ideology.

DISCUSSION

In view of the interrelationships between scales and the patterning of scale means, it is suggested that a more or less consistent psychological trend exists in the Negro middle-class. There is no clear evidence of a blind "response set to acquiesce" which Bass (3) believes can account for many correlations of the F scale; most *Ss*, even many high on F, did not "acquiesce" to the negatively formulated AW scale. The evidence does indicate that a particular kind of authoritarianism—identification with and conformity to values and norms of the white majority—is prevalent among middle-class Negroes. Like authoritarian Jews (1), they are prone to view prejudice as caused by the deviant behavior of certain members of the minority, and they are loath to make generalized criticisms of the majority group. They readily agree with the phenomenally conventional, moralistic, and ultraconservative F scale items, and their politico-economic attitudes are largely conservative.[6] It is suggested that the standards with which Negro students identify are the narrowly perceived and rigidly adhered to values of the white middle class. This is not to say that the Negro middle class is altogether antidemocratic; the sample tended to reject ethnocentric items, perhaps because they have become sensitized to overt prejudice and because Negro authorities strive to foster tolerance (5).

One interesting aspect of the results concerns those *Ss* who score high on the F scale. These *Ss* not only accept most of the anti-Negro stereotypes, but they also agree with many of the anti-white statements. Their ideological position is somewhat contradictory in that they identify with white middle-class standards, yet are critical of white people. However, these *Ss are* Negro despite their conflicts about it, and whites do not accept them as individuals. They can agree with AW items because these refer primarily to specific behaviors of whites rather than to white middle-class ideology. They implicitly deny any connection between the behavior and the ideology.

These findings may be understood, perhaps, as a partial "identification with the aggressor" (7). By accepting the respectability and propriety of white middle-class ideology and by dissociating himself from Negroes in general the Negro college student apparently attempts to deny the feeling (8) that he himself may be unworthy of acceptance by the majority.

[6] No direct evidence is available as to why the *Ss* are inconsistent in their politico-economic conservatism but we may speculate as follows: Although essentially conservative, middle-class Negroes do have an interest in the adoption of some liberal measures as a means of increasing their status. Thus, their politico-economic attitudes, though inconsistent from the standpoint of whites, are actually consistent with their general outlook.

SUMMARY

The aim of this study was to investigate authoritarian ideologies in a group of Negro *Ss*. The questionnaire method was used. Opinion-attitude scales of the Likert type were constructed to measure anti-Negro and anti-white ideologies. The California E scale (anti-Negro subscale omitted), F scale, and a revision of the California PEC scale were also used. Questionnaires were administered to a large sample of students attending colleges with primarily Negro enrollment.

Results were as follows:

1. Intercorrelations of the scales were positive and, for the most part, of moderate degree.

2. Mean scale scores indicated group tendencies to disagree with ethnocentric and anti-white ideologies, but to agree with authoritarian, anti-Negro, and conservative ideologies.

The data were interpreted to indicate an attempt by the Negro middle class to identify with stereotyped white middle-class values and to dissociate itself from other Negroes.

REFERENCES

1. ADELSON, J. B. A study of minority group authoritarianism. *J. abnorm. soc. Psychol.*, 1953, **48**, 477–485.

2. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.

3. BASS, B. M. Authoritarianism or acquiescence? *J. abnor. soc. Psychol.*, 1955, **51**, 613–623.

4. BAYTON, J. A. The racial stereotypes of Negro college students. *J. abnorm. soc. Psychol.*, 1941, **36**, 97–102.

5. BIRD, G., MONACHESI, E. D., & BURDICK, H. Infiltration and the attitudes of white and Negro parents and children. *J. abnorm. soc. Psychol.*, 1952, **47**, 688–699.

6. DOLLARD, J., DOOB, L., & MILLER, N. *Frustration and aggression*. New Haven: Yale Univer. Press, 1939.

7. FREUD, ANNA. *The ego and the mechanisms of defence*. New York: International Universities Press, 1946.

8. KARDINER, A., & OVESEY, L. *The mark of oppression*. New York: Norton, 1951.

9. MORROW, W. R. Personal communication. 1953.

10. PITT, W. A study of Negro ideological tendencies. Unpublished master's dissertation, Western Reserve Univer., 1951.

11. RADKE-YARROW, MARIAN, & LANDE, B. Personality correlates of differential reactions to minority group belonging. *J. soc. Psychol.*, 1953 **38**, 253–272.

12. SARNOFF, I. Identification with the aggressor: some personality correlates of anti-semitism among Jews. *J. Pers.*, 1951, **20**, 199–218.

Received July 20, 1956.

PERSONALITY, ANXIETY, AND MUSCLE TENSION AS DETERMINANTS OF RESPONSE STEREOTYPY

JAMES L. KUETHE AND CHARLES W. ERIKSEN[1]

The Johns Hopkins University

EXPERIMENTAL investigations of the effects of anxiety upon performance have succeeded in showing a number of ways that anxiety influences behavior. Little research emphasis, however, has been devoted to determining whether the observed performance effects are due to anxiety *per se* or whether they are characteristic of any strong drive state. Similarly, there has been little effort to determine which of the many properties of anxiety is responsible for the observed changes in psychological functioning.

Anxiety as clinically observed and defined has a number of different properties. Not only are there objectively observable autonomic nervous system indicators, but one of the most prominent symptoms of anxiety is the increase in generalized muscle tenseness or activity. In popular terminology the anxious person is "tense." Meyer (9), and Matarazzo, Ulett, and Saslow (8) have suggested that certain of the effects of anxiety upon responses may be due to the accompanying muscular tension. The plausibility of this suggestion is supported by many striking parallels between current research on anxiety effects and the older research on the effects of experimentally induced muscular tension so well summarized by Courts (2).

The purpose of the present study was to determine whether certain of the effects of anxiety upon performance could be traced to the increased muscular tension that accompanies anxiety. To accomplish this end, the stereotypy and bias with which *Ss* used a series of responses under conditions of experimentally induced muscle tension and experimentally induced anxiety were studied.

Design

It has been previously shown (5) that increased bias and stereotypy occur in the use of number responses on an absolute judgment task when the *Ss* are made anxious through random electric shocks. In the present study, a similar response effect was employed, but in order to increase the opportunity for bias to occur, the responses were freed from external stimulus determination. This was accomplished by disguising the experiment as an extrasensory perception study. *Ss* were told that *E* would concentrate on the numbers 1–11, one at a time, and they were to try to determine through ESP what number *E* was looking at. The frequency and the sequential order with which the *Ss* used the different number responses was recorded.

There were three experimental conditions. In the anxiety condition, *Ss* were randomly shocked on the ankle during the course of the experiment. In the muscle tension condition, *Ss* were required to squeeze a hand dynamometer with a predetermined force prior to and during each response. The third condition was a control where neither anxiety nor muscle tension was experimentally introduced.

As a further test of the hypothesis that the anxiety effects on this task are due to the accompanying muscular tension, *Ss* were selected on the basis of high or low scores on the psychasthenia scale from the MMPI. Previous research (3, 4) has shown that there is an interaction between high and low psychasthenia scores and performance under conditions of induced anxiety. If a similar interaction between psychasthenia extremes and muscle tension were found to occur, then the hypothesis that anxiety influences performance through increased muscle tension would be more strongly substantiated.

METHOD

All *Ss* were run individually. At the beginning of the experiment they were given a brief summary of the studies that have been done on ESP and were told that the purpose of the present experiment was to duplicate some of the findings. The *E* announced that he would draw a number from 1–11 from a pack of cards and concentrate on it; when the signal light flashed, *S* was to give the first number in this range that came to his mind. The *Ss* were shown a large pack containing numbered cards which *E* supposedly would draw from. The *Ss* were informed that the numbers 1–11 would occur in a random order, and care was taken to insure that the *Ss* understood the concept of randomness. The *Ss* assigned to the control condition were given no further instructions.

The *Ss* assigned to the muscle tension condition were further instructed that *E* was interested in the effects of muscular activity on the ESP phenomena.

[1] Now at the University of Illinois.

Subjects in this group were asked to squeeze a hand dynamometer with maximum force, using their preferred hand. One third of this force was the level of tension *S* had to maintain while responding in the experiment.

In the anxiety condition, *Ss* were told that *E* was interested in the effect of electrical stimulation upon ESP performance. They were fitted with ankle electrodes through which faradic shocks were given by means of an inductorium supplied with a 3-volt D.C. source. The *Ss* were carefully instructed that the shocks they would receive were in no way connected with their responses or their performance on the task. Prior to beginning the experiment, the maximum shock intensity that each *S* would tolerate was determined.

All *Ss*, regardless of experimental condition, received a response signal at 15-sec. intervals and were given a minute rest period at the end of each series of 12 responses. The frequent rest periods, while the same for all *Ss*, were primarily for the benefit of the muscle tension group in order to avoid their becoming fatigued. During the experiment each *S* responded a total of 154 times.

In the muscle tension group a special signal light flashed five seconds before the appearance of the response signal. When the special light flashed, *Ss* were required to squeeze the dynamometer with the predetermined force (one-third of their maximum effort) and to maintain this grip until the special light went off five seconds after the response signal had been given.

The anxiety group received 25 shocks randomly distributed through the 154 trials. Seven of these shocks were of maximum intensity for the particular *S*, 10 medium, and 8 weak. No shocks were given until 10 seconds had elasped since the last response in order to disconnect the two events as much as possible. The same random shocking order was used for all *Ss* in this group.

Apparatus

The *Ss* sat on the opposite side of a desk from *E*, separated by a wooden partition mounted on the desk. The partition contained two translucent windows which could be illuminated by *E*. A flash of light in one of the windows was the signal for *S* to respond. The other window was used only for the muscle tension group to signal when to grip the dynamometer. In order to insure that *Ss* in the muscle tension group were maintaining the required tension, the dynamometer was mounted on the desk and could be observed by *E* through an aperture in the partition.

Subjects

The *Ss* were male undergraduate students from the introductory psychology course at the Johns Hopkins University. They were required to serve a minimum number of hours as experimental *Ss* as part of the course requirements. The group form of the MMPI was administered to the entire class of 154 students, and the *Ss* used in the present experiment were obtained from those students scoring in the highest and lowest quartiles of the class on the psychasthenia scale. In each of the three experimental conditions there were 16 *Ss*, eight whose scores were in the upper quartile on psychasthenia and 8 whose scores fell in the lower quartile.

RESULTS

In order to measure the stereotypy or predictability with which *Ss* used the eleven response numbers, uncertainty statistics derived from information theory (7) were used. The basic measure, response uncertainty (U_r), is equal to $\sum p \log_2 p$, where p is the proportion of the time each response occurred. In the present study where *Ss* had available 11 different responses, response uncertainty would have had a maximum value of 3.48 bits if *S* had used each of the 11 responses equally often. The extent to which *S* departed from equal usage of all of the 11 responses is reflected in a corresponding reduction in the value of U_r. It is to be noted that failure to use all responses equally often decreases uncertainty as to what an *S*'s response will be or—alternatively stated—it increases the predictability of his behavior.

The value of U_r was computed for each *S* in each of the six subgroups. These values were then subjected to a two-way classification analysis of variance (psychasthenia groups and experimental conditions). The only significant effect obtained from this analysis was the interaction of the high and low psychasthenia groups with the three experimental conditions. ($F = 3.44$, $p < .05$ with *df* of 2 and 42). The nature of this interaction can be seen in Table 1 where the mean U_r is shown for each of the subgroups.

The table shows that *Ss* with low scores on psychasthenia are less predictable or biased in their use of the eleven response numbers under the control condition than are the *Ss* with high psychasthenia scores. However, when anxiety or muscle tension is introduced, the low group

TABLE 1

AVERAGE RESPONSE UNCERTAINTY, U_r, IN BITS AS A FUNCTION OF PSYCHASTHENIA AND EXPERIMENTAL CONDITIONS

	Low Psychasthenia Group	High Psychasthenia Group
Control	3.41	3.26
Muscular Tension	3.31	3.37
Anxiety	3.30	3.33

Using the estimate of the population variance, .0143, obtained from the error term in the analysis of variance, mean differences of .12 and .16 are significant at the .05 and the .01 levels of confidence, respectively.

becomes more predictable or biased while the high group becomes less so. If the individual cell means are compared, it is found that the mean difference between the high and the low groups in the control condition is significant beyond the .02 level of confidence, but the difference between the control and the anxiety and muscle tension conditions within the high and low groups only reach significance levels of about .10. If, however, the tension and anxiety conditions are combined in the high and low groups, respectively, and tested against the corresponding control condition, it is found that the high psychasthenia *S*s show a significant decrease in predictability or bias of responses under tension and anxiety ($p < .05$) while the low psychasthenia *S*s show a significant increase ($p < .05$). Such a combination of groups is justified since the mean differences between the anxiety and tension conditions do not begin to approach acceptable significance levels in either the high or low psychasthenic group.

This analysis of response uncertainty, U_r, deals with one way in which an *S*'s responses are predictable. It reveals the extent to which *S* demonstrates number preferences or biases in the use of the 11 response numbers. There is, however, another dimension along which an *S*'s behavior can be described as stereotyped or predictable. This is in terms of the second order or sequential dependencies among his responses. For example, an *S* may use each of the 11 response numbers equally often but persist in following his response "2" with the response "8" and response "5" with response "9." In this case, his behavior would show a fairly high degree of predictability since if one knew what his preceding response had been, one would be able to make a fair prediction of his next response.

To determine and compare the sequential dependencies in the present data, uncertainty statistics were again employed. The conditional uncertainty measure, $U_{r-1,r}$ gives the uncertainty in bits in predicting a response, r, when the preceding response, r-1, is known (7). Mathematically, $U_{r-1,r}$ is defined as $\sum p_1 \sum p_2 \log_2 p_2$, where p_1 is the probability of any given response occurring and p_2 is the probability for each response following a specified response. As in the previous analysis, maximum uncertainty would equal 3.48 bits if there were no sequential dependencies in the responses and each of the 11 responses was used equally often.

The conditional uncertainty in each *S*'s responses was computed and analyzed via a two-way classification covariance analysis (psychasthenia groups and experimental conditions). Since the value of U_r sets an upper limit upon the value that $U_{r-1,r}$ can assume, it was necessary to make a correction via the covariance method for mean differences between the subgroups on U_r. The results of this covariance analysis were essentially similar to those obtained from the analysis of the response uncertainty measures. Neither of the main effects was significant, but again there was a significant interaction between the high and low psychasthenia groups and the three experimental conditions ($F = 3.86$, $p < .05$). The nature of this interaction can be seen in Table 2 where the unadjusted mean conditional uncertainties are shown for each of the six subgroups.

TABLE 2

AVERAGE CONDITIONAL UNCERTAINTY, $U_{r-1,r}$, IN BITS AS A FUNCTION OF PSYCHASTHENIA AND EXPERIMENTAL CONDITIONS

Conditions	Groups	
	Low Psychasthenia	High Psychasthenia
Control	2.73	2.58
Muscular Tension	2.65	2.71
Anxiety	2.60	2.65

As is evident, the pattern of the interaction is the same as that obtained in the previous analysis of U_r. The low psychasthenia group shows less stereotypy in the control condition, but under tension or anxiety these *S*s show an increase in stereotypy or response predictability. The high psychasthenia *S*s show the reverse effect. In the control condition they are more stereotyped than the low *S*s, but under tension or anxiety their behavior becomes less stereotyped or predictable. Again it is to be noted that there is little difference between the anxiety and muscle tension conditions in either the high or low psychasthenia group.

DISCUSSION

The results are consistent with the hypothesis that the performance effects found with anxiety are attributable to the increases in muscle tension that accompany anxiety. Not only do both muscle tensions and anxiety have

essentially the same effects upon measures of response stereotypy, but both of these variables interact in the same way with scores on the psychasthenia scale from the MMPI. Under control conditions, *Ss* with low psychasthenia scores approximate equal usage of each of the eleven available responses much more closely than do *Ss* with high scores on the psychasthenia scale. The low *Ss* also show fewer sequential dependencies in the temporal organization of their responses. However, when either anxiety or muscle tension is experimentally introduced, the low scoring *Ss* become more biased or stereotyped in their use and organization of responses while the high scoring *Ss* become less so.

Very similar interactions between anxiety and psychasthenia scores have previously been reported. Deese *et al* (3) found that the introduction of electric shocks on a nonsense syllable learning task resulted in a decrease in the number of correct responses by the low psychasthenia *Ss* and an increase by the high *Ss*. Eriksen (4) found that high and low psychasthenia groups differed on a measure of stimulus generalization when random electric shocks were administered during the task. While the nature of the responses measured in these two studies and in the present experiment differ in terms of such variables as the amount of competition between responses and the degree of response learning or differentiation, all three experiments are consistent in showing that whatever is measured by the psychasthenia scale is a significant determiner of individual differences in performance under anxiety.

The literature on the effects of muscle tension upon performance measures also contains evidence of marked individual differences in response to induced tension. An interaction effect quite similar to that obtained in the present study is reported by Bills and Stauffacher (1) and by Stauffacher (10). The *Ss* who are poor performers under control conditions on a learning task are reported more likely to improve in performance under induced tension than *Ss* who are initially good performers. They also report that some *Ss* show a complete disruption of performance when tension is introduced.

The general research on muscle tension has suggested that there is an optimum level of muscle tension for the performance of a given task with greater or lesser degrees of tension resulting in poorer performance (2). Stauffacher (10) has made use of this general finding in proposing a theory to account for the interaction between slow and rapid learners and induced tension. Essentially, the theory holds that slow learners characteristically maintain a level of tension below the optimum while rapid learners are already operating at this optimum level. The effect of induced tension would then be to improve the performance of the slow *Ss* by moving their tension level up toward the optimum and to impair the performance of the rapid *Ss* by increasing their tension level beyond the optimum.

The results of the present experiment are consistent with Stauffacher's hypothesis if the psychasthenia scale can be assumed to measure individual differences in muscle tension. Thus, Stauffacher's rapid learners would correspond to the present high psychasthenia group and his slow learners to the low group. Such an assumption would appear justified in view of the general clinical impression that *Ss* with high psychasthenia scores are more tense than low scoring *Ss*. Also the high positive correlation between the psychasthenia scale and the Taylor scale of manifest anxiety (3, 4, 6) is consistent with such an interpretation, particularly since high scores on the anxiety scale have been found to characterize more rapid learners on simple tasks (11).

The relation of the present results to Stauffacher's findings is limited by the dissimilarity between the tasks used in the two studies, although there is a greater similarity between the task in the present experiment and the learning tasks employed by Stauffacher than is at first apparent. Previous research has shown that *Ss* do not use numbers in a purely random fashion (12). Not only are there number preferences but there are sequential relations in their response series. These response biases are in part at least unique to the individual *Ss* and may be considered to reflect differences in past learning in the relations between numbers and their use. Examined from this point of view, the present results suggest that under the control condition the high psychasthenia group shows a greater influence of this prior learning than does the low group; whereas, under the tension or anxiety conditions, the effects of this prior learning become more pronounced in the low psychasthenia group and less obvious in the high group. When stated in this manner, performance in

the present study is somewhat comparable to that reported by Stauffacher.

SUMMARY

The stereotypy and bias with which *Ss* used the number series 1–11 in a pseudo-ESP experiment was determined under three experimental conditions: (*a*) experimentally induced anxiety, (*b*) experimentally induced muscular tension, and (*c*) a control condition. Twenty-four *Ss* were selected with scores in the highest quartile on the psychasthenia scale from the MMPI and another twenty-four with scores in the lowest quartile. Eight *Ss* from each extreme group were randomly assigned to each of the three experimental conditions. The responses generated by the *Ss* on the pseudo-ESP task were analyzed for differences in frequency of usage of the different numbers and for sequential dependencies among the responses.

A significant interaction was found between the high and low psychasthenia groups and the three experimental conditions. Under the control condition, the high group was found to be more stereotyped and predictable in their use of the number responses, but under the anxiety and muscle tension conditions this relationship was reversed. Both muscular tension and anxiety had the same effect upon responses relative to the control condition and interacted in the same manner with psychasthenia scores.

It was concluded that at least some of the effects of anxiety upon responses and performance may be attributed to the increased muscular tension that accompanies the anxiety.

REFERENCES

1. BILLS, A. C., & STAUFFACHER, J. C. The influence of voluntarily induced tension on rational problem solving. *J. Psychol.*, 1937, **4**, 261–272.
2. COURTS, F. A. Relations between muscular tension and performance. *Psychol. Bull.*, 1942, **39**, 347–367.
3. DEESE, J., LAZARUS, R. S., & KEENAN, J. Anxiety, anxiety reduction, and stress in learning. *J. exp. Psychol.*, 1953, **46**, 55–60.
4. ERIKSEN, C. W. Some personality correlates of stimulus generalization under stress. *J. abnorm. soc. Psychol.*, 1954, **49**, 561–565.
5. ERIKSEN, C. W., & WECHSLER, H. Some effects of experimentally induced anxiety upon discrimination behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 458–463.
6. ERIKSEN, C. W., & DAVIDS, A. The meaning and clinical validity of the Taylor anxiety scale and the hysteria-psychasthenia scales from the MMPI. *J. abnorm. soc. Psychol.*, 1955, **50**, 135–137.
7. GARNER, W. R., & MCGILL, W. J. Relation between uncertainty, variance, and correlation analysis. *Psychometrika*, 1956, **21**, 219–228.
8. MATARAZZO, J. D., ULETT, G. A., & SASLOW, G. Human maze performance as a function of increasing levels of anxiety. *J. gen. Psychol.*, 1955, **53**, 79–95.
9. MEYER, D. R. On the interaction of simultaneous responses. *Psychol. Bull.*, 1953, **50**, 204–220.
10. STAUFFACHER, J. C. The effect of induced muscular tension upon various phases of the learning process. *J. exp. Psychol.*, 1937, **21**, 26–46.
11. TAYLOR, J. A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61–64.
12. YULE, G. U. On reading a scale. *J. Roy. Statist. Soc.*, 1927, **90**, 570–587.

Received July 21, 1956.

LEADING AND FOLLOWING PSYCHOTHERAPEUTIC TECHNIQUES WITH HYPNOTICALLY INDUCED REPRESSION AND HOSTILITY[1]

JESSE E. GORDON[2]

The Pennsylvania State University

PSYCHOTHERAPY research has in general been refractory to experimental methods. A set of procedures is needed by means of which laboratory research into the influence of a variety of patient, therapist, and outcome variables may be conducted. These procedures, however, must be such that results may be generalized to clinical practice. The ecological validity required of such research makes it necessary that subjects be permitted a wide range of responses, as are patients in therapy. Such requirements make experimental control and direct behavioral measurement difficult.

These problems are not so acute when the experimental variables are global ones such as directive vs. nondirective therapy, therapeutic success as defined by various criteria, and patient "maladjustment." However, systematic investigation aimed at building up a body of knowledge regarding therapeutic practices, instead of simply "proving" one school of thought to be superior to another, requires more refined control and manipulation of discrete, unitary variables rather than global ones. When a body of knowledge has been acquired in which the relationships among particular therapeutic practices and specific responses by patients are known, then a therapist may decide what effect he wishes to have on a particular patient and select his therapeutic techniques accordingly. Although this state of affairs may be an unattainable goal, it seems to be the only possible ideal at which the scientific therapist can aim.

With these considerations in mind, a set of experimental procedures was developed which seemed to possess ecological validity, which permitted a measure of control over the major client, therapist, and outcome variables, and which permitted the use of direct behavioral measures as well as psychometric methods. The experimental procedures involve the use of hypnosis to make normal subjects believe and act as if they had undergone a standard, mildly upsetting experience. The *Ss* have freedom to react to the experience in their own idiosyncratic ways. Reviews by Weitzenhoffer (9), Hull (5), and Rapaport (8) present much direct and indirect evidence to support the validity of generalizing from such hypnotically stimulated reactions to reactions actually encountered in clinical patients.

The use of a standard mildly upsetting experience, hypnotically induced, puts control over an important patient variable into the hands of the experimenter. This kind of control is practically and theoretically impossible in field research in psychotherapy. The acquisition of such control by the experimenter represents the major advantage of this type of research over therapy research conducted in the clinic with a patient population.

Instructions to the hypnotized *Ss* were designed to make them believe that they had been bumped into by a physical education teacher in high school and unjustly blamed for the incident. These instructions were patterned after those used successfully by Counts and Mensh (1). In addition, it was suggested to *S* that the therapist who was to interview him after he awakened bore some resemblance to the physical education teacher. Posthypnotic amnesia for the upsetting experience was also induced. These hypnotic instructions are based on a neurosis model which characterizes neurotically inappropriate behavior as resulting from an unconscious identification of the patient's present situation with an unpleasant and often unremembered experience in the patient's past.

The independent variables of this experiment were therapist's Leading (asking questions, making suggestions, etc.) and the therapist's Following (restating, reflecting, clarifying). These two kinds of therapist verbal behavior were selected for experimental comparison on the basis of investigations which indicate that they reflect a basic split among schools of psychotherapy (2, 4, 7). Roughly

[1] This article is based upon a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the Pennsylvania State University. The writer wishes to acknowledge his indebtedness to Dr. George M. Guthrie and Dr. William U. Snyder.

[2] Now at the University of Wisconsin.

paralleling this split regarding the use of Leading and Following behaviors is a split in what is regarded as the important outcomes of therapy. Those who use Leading techniques tend to regard the lifting of repressions as the aim of therapy. Those who use Following techniques generally disclaim interest in the historical approach to therapy and focus instead on improvement in the patient's present interpersonal functioning as the aim of therapy. It seems appropriate, therefore, to employ measures of both types of outcome variables, repression and interpersonal functioning. The hypnotic instructions to forget the "traumatic" experience produce behavior analogous to repression, making possible the measurement of the lifting of repression. The inclusion in the hypnotic instructions of a suggested similarity between the teacher in the traumatic experience and the therapist provides a situation analogous to transference, and thus makes possible some measurement of the inappropriate, "neurotic" interpersonal behavior of *S*.

The specific purpose of the present research, then, was to test a set of procedures for the experimental investigation of psychotherapy by applying them to a study of the effects of therapist Leading and therapist Following on hypnotically induced repression and on neurotic interpersonal functioning.

METHOD

Procedure

Sixty-two male college students volunteered to participate in "an experiment on hypnosis." Two group and then one individual training and screening sessions were held, out of which 26 of the volunteers emerged as susceptible enough to obey a hypnotic suggestion to be amnesic for trance events. Of this group, the data on eight subjects were lost as a result of faulty recordings of the interviews. Thus, 18 *S*s remained in the two experimental conditions, Leading and Following.

The *S*s were appointed for the individual experimental sessions following the third training and screening session. During this third session they were instructed, while hypnotized, to imagine themselves in a situation in high school in which a physical education teacher accidently bumped into the *S* in a hallway, then blamed *S* for clumsiness and threatened reprisal. The *S* was instructed not to remember the incident after he was awakened and was told to have a daydream about it upon a signal. This served as a check on the acceptance of the hypnotic suggestions and as practice in covertly playing the role of an unjustly blamed person. Following this test, the graduate student "therapist" to whom *S* had been assigned was introduced to the subject. The *S* was then rehypnotized and told that the incident with the physical education teacher had actually occurred. In addition, it was specified that the physical education teacher looked "something like Mr. ———, the man who will be interviewing you soon." The instructions used to induce belief that the incident had actually occurred were those used with success by Erickson (3). In addition to Erickson's instructions, it was further specified that though *S* would not be able to recall the experience when he awoke, it would bother him, and he could spend the interview hour trying to figure out what was bothering him. The instructions made it permissible to recall the experience during the course of the interview. Before being awakened, *S*s were made amnesic for the entire session, including the introduction of the therapist.

The *S* was then awakened and brought into the interview room for a 50-minute interview, which was tape recorded. Following the interview, *S* was again rehypnotized, permitted to recall the content of the previous hypnosis, and disabused of the experience.

The therapists were nine advanced graduate students in clinical psychology at the Pennsylvania State University. Each therapist was in possession of a statement describing the acceptable therapist verbal behaviors in the Leading and Following conditions. These were explained and discussed more fully in a group meeting with the therapists. However, the therapists were not told the nature of the dependent variables or that a paramnesia had been induced in the *S*s. Each therapist appeared twice, once in each condition, in an effort to control intertherapist differences. The order in which each therapist appeared in the conditions was counterbalanced.

Typescripts were prepared from the tape recordings of the interviews. Therapist statements on all even-numbered typescript pages were categorized by the *E* as Leading, Following, and Miscellaneous as a check upon the extent to which the experimental conditions were realized.

Repression. The content of the experimental "traumatic experience" had been broken down into meaning units, and criteria for scoring recall of the units were prepared. Three independent graduate student judges scored the typescripts, using these criteria to obtain scores for an Amount Recalled measure of the extent to which repression had been lifted. The means of the recall scores assigned by the judges constituted the scores on this variable. Agreement among the judges was high (r = .94, .98, .93).

The judges were also asked to indicate the statement in each typescript containing each *S*'s third scorable recall item. By locating these *S* recall statements on the tape recordings, it was possible to record the amount of elapsed time from the beginning of the interview, thus yielding a Recall Time score of the speed with which successful lifting of repression began to occur.

One other measure related to the lifting of repression was designated Recognition Time. To obtain this measure, the judges were asked to check the first *S* statement in each typescript which contained a verbalized awareness that the therapist, or someone like him, was familiar to *S*. Again, by locating the checked *S* statement on the tape recordings, it was possible to measure the elapsed time from the beginning of the

interview to the point at which this kind of insight began to develop.

Hostility. Hostility related behavior of the *S*s was measured by means of rankings. Verbal descriptions of three aspects of hostility were given to three graduate students in a course in therapeutic interviewing at the University of Wisconsin, who then ranked the typescripts on each aspect. These hostility variables were (*a*) Extrapunitiveness, in which *S* attempted to convey his hostility to the interviewer in a direct manner, (*b*) Passive Hostility, in which hostility was expressed through negativism, uncooperativeness during the interview, and general resistance to the therapist, and (*c*) Disguised Hostility, in which *S*'s hostility was communicated in symbolized or other indirect fashion, such as discussion of hostile dreams, overemphasis on aggressive feelings toward others, etc. Coefficients of concordance revealed significant agreement among the judges in their rankings on the first two variables but not on Disguised Hostility, which was therefore not included in further statistical analyses. Each *S*'s rank on the two hostility variables was the mean of the ranks assigned to him by the three judges. Kendall's tau was computed between the rankings on Extrapunitiveness and on Passive Hostility and was found to be nonsignificant. The two indexes were therefore regarded as independent.

RESULTS

The results of the classification of the therapists' verbal behaviors are presented in Table 1 in terms of frequencies and proportions.

TABLE 1

MEAN FREQUENCY AND PROPORTION OF THERAPIST LEADING AND FOLLOWING STATEMENTS IN THE LEADING AND FOLLOWING CONDITIONS

Measure	Leading Condition		Following Condition	
	Leading	Following	Leading	Following
Frequency:				
Mean	26.88	10.22	10.55	27.22
Sigma	12.22	8.48	10.93	13.64
Proportion:				
Mean	.64	.21	.20	.65
Sigma	.21	.16	.14	.19

Repression

Amount Recalled. Significance testing was in terms of the differences between each therapist's Leading and Following interviews. Six differences showed greater Amount Recalled in the Leading condition with two differences favoring the Following condition and one of zero value. A randomization test of the differences, described by Moses (6), indicates that this distribution of differences just falls short of significance at the 5 per cent level with a two-tailed test.

Recall Time. Analysis of Recall Time was again in terms of the differences between each therapist's Leading and Following interviews. Three differences showed faster recall in the Leading condition; five differences favored the Following condition, and one difference was zero. A randomization test of the differences indicates that this distribution does not depart significantly from chance ($p > .05$).

Recognition Time. The differences in Recognition Time between each therapist's Leading and Following interviews indicated more rapid recognition in the Leading condition five times, three differences favoring the Following condition, and one difference of zero. A randomization test of the differences indicates that this distribution does not depart significantly from chance ($p > .05$).

The correlation between Amount Recalled and Recall Time was $-.78$, between Amount Recalled and Recognition Time $-.55$, and $+.50$ between Recall Time and Recognition Time. All three correlations are significant at the .05 level. The very high correlation between Amount Recalled and Recall Time probably represents the extent to which these variables measure *S*'s ability to repress and to conform to the hypnotic suggestions. On this assumption, an analysis of variance of errors of estimate in the regression of Amount Recalled on Recall Time was carried out. In effect, this procedure partialed out intersubject variability in conformity to the hypnotic suggestion. The results of this analysis are presented in Table 2[3].

[3] It is important to recognize that this analysis is not an analysis of covariance. In this analysis, the matching variable (Recall Time) was obtained under the experimental conditions and was presumably affected by the conditions. In covariance analysis, the matching variable must be measured before the experimental conditions are imposed. The analysis of variance of errors of estimate as carried out here is, in effect, a test of the significance of the difference between the correlation between Amount Recalled and Recall Time in the Leading condition and in the Following condition. The analysis demonstrates that the correlation is very high in the Leading condition ($-.93$), thus the interpretation that *S*s are recalling to the limits of their abilities in this condition; it is much lower in the Following condition ($-.63$), thus yielding larger errors of estimate and indicating that other factors have influenced the Amount Recalled, reducing it below the limits of the *S*'s ability to recall as this ability is estimated from the Recall Time variable.

TABLE 2
ANALYSIS OF VARIANCE OF ERRORS OF ESTIMATE IN THE REGRESSION OF AMOUNT RECALLED ON RECALL TIME

Source	*df*	Mean Square	*F*
Between Treatment	1	62.94	12.87**
Among Therapists	8	17.66	3.61*
Residual	8	4.89	
Total	17		

* Significant at .05 level.
** Significant at .01 level.

Table 2 indicates that the errors of estimate in the regression of Amount Recalled on Recall Time were significantly larger in the Following condition when differences among therapists are eliminated from the comparison. It would appear that in the Leading condition, *S*s tended to verbalize the "repressed" material more nearly to the limits of their ability to recall; whereas in the Following condition, other variables beyond the basic recall factors entered into the determination of how much recalled material was verbalized, thus producing larger errors of estimate and less verbalized recall. Differences among therapists were also found to be significant, which probably reflects differences in therapeutic skill.

The interviews were divided into those which were the therapists' first interviews and those which were second, and then analyzed in order to test the effect of therapist experience in this sense. Randomization tests of the differences indicate that first interviews produced significantly more recall than second interviews ($p < .05$), but Recall Time and Recognition Time variables were not significantly different from first to second interviews.

All interviews were then combined and ranked in terms of the number of Leading statements, the proportion of Leading statements, and on the three repression variables. Kendall's tau statistic was computed between all pairs of independent and dependent variables. The only relationship found that approximated significance was expressed by a tau coefficient of .31, with an associated probability of .075, between Amount Recalled and proportion of Leading statements.

Hostility

Mann-Whitney *U* tests indicate that there was no significant tendency for Leading or Following interview *S*s to be ranked higher on either hostility variable. Chi square was computed between classification of the interviews as above or below the median rank on Extrapunitiveness and their classification in the same terms on Proportion Leading. The value obtained after correction for continuity was 3.56, just short of significance at the .05 level. Similar analyses between Extrapunitiveness and Amount Leading, and between Passive Hostility and Amount and Proportion Leading, yielded definitely nonsignificant results.

The rank order correlation between rankings on Extrapunitiveness and on Recognition Time was .44, which is significant beyond the .05 level. Passive Hostility was not significantly correlated with Recognition Time, and neither of the two hostility variables was found to be significantly related to the other repression variables. The relationship between Extrapunitiveness and Recognition Time would be expected on the basis of the nature of the experimental instructions regarding the similarity between the therapist and the teacher. An *S* who quickly "recognized" his interviewer would be expected to be, and was, more hostile toward him than one whose recognition was delayed. One cannot decide, on the basis of these data, whether the recognition led to the hostile expression or *vice versa*.

As in the analysis of the repression variables, the typescripts were divided into those for the therapists' first interviews and those for their second, and analyzed for any relationships between sequence and hostility. A Mann-Whitney *U* test yielded a probability of less than .01, indicating that *S*s were significantly more extrapunitive in the therapists' second interviews than in their first. It will be remembered that there was also significantly less recall in these interviews.

DISCUSSION

The results of this experiment indicate that Leading therapist behaviors tended in the samples used to be more efficient in obtaining as much verbalization of "repressed" material as *S* was capable of. A review of the typescripts lends support to this interpretation; in the Following interviews one notes several instances in which *S*s failed to pursue what appeared to be a productive line of thought, ostensibly because of the lack of stimulation from therapist.

The finding that therapists' first interviews were more productive than second interviews,

in terms of the amount of recall, appears contrary to what one might expect. Second interviews should have reflected the insights gained by the therapists in their first interviews, and therefore might have been expected to be more productive. A suggested explanation for the findings is that the poorer results in the second interviews arose as a result of generally incorrect hypotheses developed by the therapists on the basis of their experiences in the first interviews. None of the therapists were able correctly to identify the nature of the experiment or the role of hypnosis after the first interview. Considering the many things about which *Ss* spoke in the interviews, it was highly probable that therapists would give undue weight to irrelevant verbalizations in forming hypotheses about the experiment and the *Ss*. This may be a function of the relative lack of experience of the graduate students who served as therapists. If this interpretation is valid, the danger of premature diagnostic hypotheses becomes apparent.

The significantly greater Extrapunitiveness of *Ss* in the therapists' second interviews is of interest. The relationship between Extrapunitiveness and Recognition Time indicates that the hostility is at least in part a product of the experimentally induced negative transference. The data do not permit a clear understanding of the relationship between this negative transference and the serial position of the interviews. It will be recalled that *Ss* in the therapists' second interviews had significantly lower Amount Recalled scores. Possibly the direct hostility in these interviews inhibited the recall, but it appears equally plausible that the unresolved repressions in the second interviews produced tensions in *Ss* expressed as Extrapunitiveness. Neither of these hypotheses was supported by the nonsignificant correlations between Amount Recalled and Extrapunitiveness. However, the experimental design used in the present research provides a method for testing these hypotheses more systematically in the future.

Many of these suggested conclusions stem from inferences drawn from relatively unreliable data. However, the experimental technique employed in the present study has demonstrated its usefulness in studying psychotherapy variables and generating new hypotheses. That these procedures bear striking similarities to clinical practice is demonstrated by the behaviors of the *Ss*, who appeared genuinely anxious about the "repression." They seemed highly motivated yet natural in their associations, defenses, topics of conversation, etc.

Of incidental interest in reviewing the typescripts are the chains of associations by which *Ss* approached and eventually resolved the repression. These associations were natural, involving the telling of past experiences in the *Ss*' lives related to the "repressed" incident. For example, one *S* began the interview by describing his hostility and guilt when, as a small boy, his mother had punished him unjustifiably. She later overheard him expressing his anger toward her to a playmate. The relevance of this memory to the induced experience is obvious. It is interesting that this *S* eventually recalled most of the induced experience but could not correctly identify the relationship between the therapist and the physical education teacher. However, he did manufacture a plausible though incorrect reason for his feelings toward the therapist and experienced a real sense of relief when he hit upon this "reason." This observation suggests that true insight may not be a necessary prerequisite for relief from anxiety. It is quite possible that reasonable rather than accurate schematizations of experience contribute most to therapeutic effectiveness. Other interesting phenomena encountered in the interviews were dynamic determiners in word choice ("I'm sure we've ***bumped into*** each other somewhere or other") and the many ways in which hostility toward the therapists was expressed.

Other examples of realism in the interviews are provided by one *S* who spent the entire interview time "spontaneously" discussing difficulties in controlling frequent and intense anger. Another *S* experienced much embarrassment in relation to the therapist, whom he eventually discovered reminded him of his high school physical education teacher. He described the teacher as a father figure who disappointed him by not always behaving in the perfect way in which the *S* idealized him. The specific incident which *S* finally described as support for his disappointment was the induced one. Thus, there is considerable clinical evidence that the reactions of the *Ss* to the induced experience were idiosyncratic and clinically realistic.

The procedures used here appear to be practicable, and they potentially provide reliable and quantifiable information on a number of important psychotherapeutic issues: the rela-

tionships between the lifting of repressions and transference and hostility, the effects of therapist experience, the interaction of therapist and client personalities, the effects of therapist's sex on male and female *S*s, etc. The procedures in their present form, however, are not without their limitations. The use of hypnotically susceptible *S*s places an important restriction on the extent to which experimental results may be generalized. The restriction of this kind of study to a single interview with each *S* makes it difficult to study such procedures as the free association method, which requires fairly extensive training of *S*. Further, the traumatic neurosis model implied in the experimental design does not take into account the observation that neuroses occur only as a result of many such experiences in persons of particular personality structures.

Finally, there are problems in controlling therapist behavior. It is fundamental to the present study that the independent variable is a unitary, easily defined verbal response of the interviewer. An attempt was made to select for study a verbal response similar to those frequently made during psychotherapy. However, psychotherapy cannot be defined strictly by such limited verbal behaviors. The purpose of the present research was *not* to study psychotherapy; rather it was to provide a framework for studying variables involved in psychotherapy. This is an important distinction, and one which makes it necessary to select meaningful but well-defined variables while controlling as many associated contaminating variables as possible. As long as experimental research is concerned with different kinds of therapy, it will suffer from such problems as those involved in having therapists artificially change their professional styles, difficulties related to the experience of therapists, etc. Such problems may be largely avoided when research is conceived as chiefly concerned with particular classes of verbal response in face-to-face interactions.

These considerations emphasize the apparently essential limitations of experimental research in psychotherapy in that complexity and richness may be sacrificed in order to obtain reliable answers. However, there is still much to be profitably learned about the kinds of variables included within the relatively wide limitations of these experimental procedures. Inasmuch as this method makes possible the reliable measurement of fairly discrete variables, it has value for current investigators. It may be important to find answers for the questions that can be confidently answered at present before going on to find less reliable answers to more difficult questions.

SUMMARY

A set of experimental procedures for use in psychotherapy research was tested. These procedures involved the use of hypnosis for the induction of miniature neuroses in normal individuals. Therapist Leading and Following, operationally defined, were studied for their effects on hostility and on the lifting of repressions in one-hour therapeutic interviews.

The findings indicate that Therapist Leading is more effective than Following for lifting repressions. It was also found that therapist experience which might lead to the formation of incorrect hypotheses about a client has a deleterious effect on the lifting of repression and is also associated with stronger expression of transference hostility by the client. The usefulness and limitations of the experimental procedures in providing possibilities for better control and measurement of psychotherapy variables were emphasized.

REFERENCES

1. COUNTS, R. M., & MENSH, I. N. Personality characteristics in hypnotically-induced hostility. *J. clin. Psychol.*, 1950, **6**, 325–330.
2. DANSKIN, D. G., & ROBINSON, F. P. Differences in "degree of lead" among experienced counselors. *J. counsel. Psychol.*, 1954, **1**, 78–83.
3. ERICKSON, M. H. The method employed to formulate a complex story for the induction of an experimental neurosis in a hypnotic subject. *J. gen. Psychol.*, 1944, **31**, 67–84.
4. GUMP, P. V. A statistical investigation of one psychoanalytic approach and a comparison of it with nondirective therapy. Unpublished master's dissertation, Ohio State Univer., 1944.
5. HULL, C. L. *Hypnosis and suggestibility*. New York: Appleton-Century, 1933.
6. MOSES, L. E. Non-parametric statistics for psychological research. *Psychol. Bull.*, 1952, **49**, 122–143.
7. PORTER, E. H. The development and evaluation of a measure of counseling interview procedures. II. The evaluation. *Educ. Psychol. Meas.*, 1943, **3**, 321–350.
8. RAPAPORT, D. *The organisation and pathology of thought*. New York: Columbia Univer. Press, 1951.
9. WEITZENHOFFER, A. *Hypnotism: An objective study in suggestibility*. New York: Wiley & Sons, 1953.

Received July 27, 1956.

THE EFFECTS OF PERSONAL AND SHARED THREATS UPON SOCIAL PREJUDICE[1, 2]

SEYMOUR FESHBACH AND ROBERT SINGER

University of Pennsylvania

THE effectiveness of anxiety evoking communications in modifying attitudes has been explored in a number of investigations (3, 4, 8). Janis and Feshbach (3) were particularly concerned with demonstrating that communications which attempt to modify attitudes by means of threat may elicit defensive responses that reduce the intended effect of the communication. Clinical evidence (6, 9) suggests that, in addition to avoidant behavior, threatening situations frequently elicit hostility. The purpose of the present experiment is to determine the conditions under which threatening situations elicit hostility and affect the expression of social prejudice.

The immediate stimulus for this experiment was a study carried out primarily for pedagogical purposes during a seminar on group relations conducted by the senior author. The study was designed to test the simple hypothesis that the arousal of anxiety results in an increase in social prejudice. It was presumed that the anxiety would arouse hostility, manifested in turn in the expression of social prejudice. The similarity of this formulation to the application of the frustration-aggression hypothesis to social prejudice is apparent. Whereas in a study of social prejudice such as that of Miller and Bugelski (7), frustration was varied in order to produce variation in aggression, this demonstration involved manipulation of the degree of threat present in a stimulus situation. Schematically, the two respective formulations may be compared as follows: 1. Potentially frustrating situation: frustration—aggression—increase in social prejudice; 2. Potentially threatening situation: fear—aggression—increase in social prejudice.

The manipulation of threat was accomplished by varying the content of an interview conducted by the students in the seminar. Each student interviewed approximately six white adult respondents from the community such that half of his respondent sample were randomly assigned to the Control (neutral) interview and half to the Experimental (fear-arousing) interview. In the Control interview, the respondents were asked to give their opinions on the problem of the teacher shortage and overcrowding in the classroom. Their views were also requested on the issue of "pay television." They were then asked to complete a social prejudice questionnaire consisting of items selected from the California E Scale (1) which dealt primarily with attitudes towards Negroes; the questionnaire also included several general ethnocentric items from the E scale. A similar procedure was followed for the Experimental group with the important exception that questions about cancer and about the effects of an A-bomb attack and atomic radiation were used instead of the presumably neutral items asked in the Control interview. These questions were designed to arouse anxiety and constituted the experimental variable of the study. It was hypothesized that the Experimental subjects would respond in a significantly more prejudiced direction on the questionnaire than would the Control group.

Analysis of the experimental results revealed a significant difference between the Experimental and Control groups, but in a direction opposite to that predicted. Nineteen of the 27 subjects in the Experimental group obtained scores which fell below the median of the combined distribution while the scores of only 7 of the 25 Control subjects fell below the median. This difference in proportions yields a chi square of 9.3 which is significant at the .01 level of confidence.

This unanticipated result indicates the operation of a mechanism quite different from that originally postulated. Not only did the threat-arousal interview fail to produce an increase in social prejudice, but it resulted in an actual decrement in the expression of social prejudice. It is as if the threat interview evoked sympathy or a sentiment of "brothers under the skin" in its influence on the attitudes

[1] This study was supported by a grant from the Albert M. Greenfield Center for Human Relations of the University of Pennsylvania. The authors especially wish to express their gratitude to Dr. Martin P. Chworowsky, Director of the Center, who read the manuscript and made a number of helpful suggestions.

[2] A paper based upon this study was presented at the 1957 Eastern Psychological Association meetings.

of the white respondents toward Negroes. One could, of course, relinquish the threat-aggression hypothesis. However, an alternative possibility is proposed; namely, that the type of threat is the critical parameter. Specifically, it is postulated that a personal or ego threat results in an increase in hostility and in social prejudice while a "shared" threat results in a decrement in the expression of social prejudice.

The following criteria are offered for distinguishing between a personal and shared threat with respect to their possible influence upon social prejudice. A personal threat is one which is perceived by a subject as primarily directed toward himself rather than toward other individuals—particularly toward members of other ethnic and racial groups. Threats which elicit feelings of inadequacy and loss of status are likely to be personal threats. The possibilities of loss of job and bodily injury also fall within the category of personal threats if such threats are not shared by members of other social groups. Shared threats, on the other hand, are threats which are perceived as equally affecting members of both majority and minority groups. Shared threats do not discriminate between Negro and white, Jew or Christian. All groups share not only the danger but also the responsibility of coping with that danger. The prospect of war, hurricane, or flood are situations that are likely to have the characteristics of a shared threat. This is not to say that the prospect of war or hurricane must necessarily have the characteristics of a shared threat. Some individuals, particularly those with strong egocentric or paranoid tendencies, might interpret any impending disaster in very personal terms. Similarly, a threat that most individuals may perceive as personally directed may be reacted to by others in terms of a broader frame of reference. The examples of personal and shared threats cited above are intended as illustrations of the kind of situations that are likely to have personal or shared threat properties for most respondents in the general population.

Under the impact of a common threat, it seems reasonable to assume that one's reference group may become the population that is subjected to the danger. If this reference group now includes both Negro and white, whereas under ordinary stimulus conditions the reference group has been primarily the white population, then the social distance between white and Negro should decrease, with a corresponding decrement in social prejudice. Social prejudice is at least, in part, a function of differences in group membership and such prejudice should be reduced to the extent that individuals perceive themselves as members of the same reference group; that is, as more alike due to the sharing of a significant common experience. One is reminded in this connection of the Gestalt notion of "common fate" as one of the factors determining perceptual organization.

METHOD

Materials

The type of threat was manipulated by varying the content of a mimeographed communication presented to the *Ss*. Each communication, whether Control, Personal Threat-Arousal or Shared Threat-Arousal, consisted of a brief discussion (approximately one-half page of single-spaced typing) of some issue. After reading this discussion, *S* answered a series of questions dealing with his attitudes and feelings regarding the issue.

The Control communication consisted of a discussion of pay television and of the length, frequency, and taste of television commercials. Three different Personal Threat-Arousal communications were used. One dealt with the frequency of divorce and the likelihood of making a satisfactory marital adjustment. The *S* was required to answer questions regarding the probability of his making a successful marriage and was asked to discuss any traits he had that might interfere with an adequate marital adjustment. A second Personal Threat dealt with the frequency of mental disorder and the respondent was asked to discuss his chances of becoming mentally ill. The third Personal Threat dealt with the dangers of fire and accident and the respondent indicated how he would react in the event of severe personal injury. Of the two Shared Threat-Arousal communications, one dealt with the danger of flood and hurricane while the second discussed the possibility of an atomic war.[3] In both instances, the respondent was required to indicate how he would behave if actually exposed to such disasters. The questions asked of the Mental Health Threat group and of the Flood and Hurricane Threat group are presented in detail in order to illustrate the difference between the operations used to induce a personal threat and the operations used to induce a shared threat.

Mental Health Threat questions: 1. Have you ever thought of the possibility that you may develop some form of mental illness? (Discuss). 2. In case of mental illness, to what extent can you rely on your family for financial support and sympathetic understanding? (Discuss). 3. If you developed some form of mental disease and later recovered from it, do you feel that your occupational and social adjustment would be made more difficult due to once having been mentally ill? (Discuss briefly). 4. If you became psychotic would you prefer to be treated at home or at a mental hospital? (Discuss briefly).

[3] Copies of the communications may be obtained on request from the authors.

Flood and Hurricane Threat questions: 1. About how many people do you think were killed as a result of floods and hurricanes in the past 12 months? 2. What would you do if the area you lived in were flooded or threatened by a hurricane? (Discuss). 3. What could you do if the people in your area, including close friends, couldn't get out of the danger area and all immediate help was cut off? 4. What do you think should be done for the people who would suffer great personal loss, physically, emotionally and financially, through an epidemic resulting from floods? 5. What do you think that people in your area can do to protect themselves against epidemics like the recent one in Lancaster?

The measure of social prejudice used in this study consisted of the six items from the California E scale dealing with attitudes toward Negroes, eight general ethnocentrism items, also taken from the E scale (1, p. 142) and 14 buffer items. Six alternatives ranging from "agree very strongly" to "disagree very strongly" were available for each item. The scoring of each item ranged from +3 for the most affirmative response to −3 for the most negative response. E-scale items are so constructed that agreement with an item is indicative of social prejudice.

Procedure

College students selected from introductory psychology classes served as subjects, and were assigned at random to the Control treatment, to one of the three Personal Threat treatments, or to one of the two Shared Threat treatment groups.

The social prejudice questionnaire was administered to the entire sample. Approximately four weeks later, the mimeographed communications were distributed to the *Ss* following a regular laboratory section meeting. Each laboratory section consisted of about 25 students. Because the experimental communications were silently read by each respondent, it was possible to distribute at random all communication forms to each laboratory section. Any single respondent was exposed, of course, to only one communication. After the *Ss* completed their answers to the communication, the social prejudice questionnaire was administered for the second time. Completed social prejudice questionnaires administered before and after the experimental treatments were obtained from 45 Control *Ss*, 37 Marriage Threat *Ss*, 24 Mental Health Threat *Ss*, 30 Fire and Accident Threat *Ss*, 25 Flood and Hurricane Threat *Ss* and 26 Atomic War Threat *Ss*.

It was predicted that each of the Personal Threat groups would show a significant *increase* in social prejudice in comparison to the Control group while each of the Shared Threat groups would show a significant *decrease* in social prejudice in comparison to the Control group.

RESULTS

Examination of the initial social prejudice scores indicated that the various treatment groups were homogeneous with respect to initial prejudice. Prejudice change scores were then calculated for each individual so that a positive shift reflects an increase in social prejudice, and a negative shift, a decrease. The mean and variance of the prejudice change scores for each of the experimental treatment groups are presented in Table 1. Except for the Atomic War Threat group, all changes are in the predicted direction. The difference between the Fire and Accident Threat group and the Control group is significant at the .05 level of confidence while the difference between the Mental Health Threat Group and the Control Group is significant at the .06 level. However, there is a significant difference in variance between the Marriage Threat Group and the Control Group which renders the *t* test inapplicable in this instance. The number of cases in each group falling above and below the median of the combined distribution was then determined. Twenty-four of the 37 *Ss* in the Marriage Threat group in comparison to 18 of the 45 *Ss* in the control group obtained scores falling above the median. This difference in proportions yields a chi square of 4.94, significant at the .05 level of confidence. The results for the Marriage Threat group then, in conjunction with the Mental Health Threat and Fire and Accident Threat data, lend strong support to the hypothesis that communications that arouse personal threats result in a significant increase in social prejudice.

TABLE 1

COMPARISON OF MEAN IN PREJUDICE CHANGE SCORES IN EACH THREAT GROUP VS. THE CONTROL GROUP

Treatment Group	M	σ	σ_M	t	p
Control (N = 45)	+0.33	5.09	.76	—	—
Marriage Threat (N = 37)	+3.30	7.4	1.2	2.11	<.05
Mental Health Threat (N = 24)	+3.04	6.1	1.2	1.92	<.06
Fire Threat (N = 30)	+3.03	6.2	1.1	2.10	<.05
Flood Threat (N = 25)	−1.64	5.8	1.1	1.51	>.10
Atomic War Threat (N = 26)	+2.60	7.3	1.5	1.38	>.10

In regard to the data bearing upon the effects of Shared Threats, the *t* test does not permit one to reject the null hypothesis of no difference between the Flood and Hurricane Threat group and the Control group. However, inspection of the data revealed a much greater proportion of decrements in prejudice in the Flood and Hurricane Threat group than in the Control group. The median of the combined distribution is 0, or no change, and 17 of the 25 *Ss* in the Flood and Hurricane Threat group in comparison to 18 of the 45 *Ss* in the Control group have prejudice change scores below the median, the difference yielding a chi square of 5.0, significant at the .02 level of confidence.

Thus the data for the Flood and Hurricane Threat group lend support to the hypothesis that the arousal of a Shared Threat results in a significant decrease in social prejudice. The data for the Atomic War Threat communication, however, fail to support the Shared Threat hypothesis. Because of a significant difference in variance between the Atomic War Threat group and the Control group, the number of cases falling above and below the median of the combined distribution was determined. Ten of the 26 *Ss* in the Atomic War Threat group and 26 of the 45 *Ss* in the Control group obtained scores falling below the median. While the difference is still in a direction opposite to that predicted, it is not statistically significant, chi square being 2.48.

DISCUSSION

The hypothesis that the arousal of a personal threat results in a significant increase in the expression of social prejudice is strongly supported by the data. The fact that each of three quite different communications yielded results consonant with the hypothesis provides additional confidence in the genuineness of the phenomenon. These results are readily accounted for by the theoretical analysis originally proposed; namely, the personally threatening condition arouses hostility which is then displaced on members of an out-group. While the data do not bear directly upon this issue, it is possible that the displacement of hostility in the form of increased prejudice has some functional significance in the sense of reducing the respondent's hostility or his insecurity. Thus the disparagement of the Negro implied by the increase in prejudice may partially compensate for the feelings of inadequacy aroused by the personally threatening communications.

There are undoubtedly a number of parameters related to the postulated threat-hostility-displacement sequence that warrant further study. The data indicate that hostility resulting from a personal threat is displaced on an out-group despite the lack of any conceivable connection between the out-group and the source of the threat. One may raise the question, however, as to whether it is necessary for some prejudice to be already present toward an out-group in order for such displacement to take place. Also, there are probably individual differences in response to personal threats such that some subjects may tend to direct hostility toward themselves rather than toward other individuals or groups. It would be worthwhile to explore such personality variables as extrapunitiveness and intrapunitiveness in relation to their influence on the displacement of hostility. In addition, it may be useful to study the effects of different types of personal threats. Maslow's distinction between deprivation frustration and frustration involving psychological threat (5) suggests that ego threats may result in more hostility than deprivation threats although both types of situations fall within the category of personal threats as defined in this paper.

The findings for the Shared Threat-Arousal communications, while less conclusive than those with respect to the Personal Threat-Arousal Communications, are consistent with the hypothesis that exposure to shared threats results in a significant reduction in social prejudice. Of the two Shared Threat groups, the one exposed to the Flood and Hurricane Threat showed a reliable decrease in social prejudice while the group exposed to the Atomic War Threat did not show any significant change in comparison to the Control group. There are a number of possible explanations that may account for the failure of the group exposed to the Atomic War Threat condition to behave as predicted. First, the communication may not have been particularly credible or threatening for a college population. There is evidence, however, that it did have some effect; namely, it resulted in a significantly greater variance in the prejudice change scores in comparison to those of the Control group. Some subjects responded to the Atomic War Threat with relatively large increments in prejudice while others responded with relatively large decrements in prejudice. Thus the mean increment for those in the Atomic War Threat group who increased in prejudice is +6.4 and the mean decrement for those who decreased in prejudice is −6.3, while the corresponding figures for the Control group are +4.2 and −5.4.

While the atomic threat was effective in reducing prejudice in the original study, there are important differences between the threat-arousing communications employed in the two studies. In the original study, the effects of the atomic threat were confounded with reactions

to the cancer threat.[4] In the present study, the atomic threat was directly linked to a possible war with the Soviet Union. Because of this aspect of the communication, the responses to the general ethnocentrism items and the responses to the Negro items were separately analyzed. However, score changes in the Atomic War Threat group were similar on both sets of items. This correlation also held for the other threat groups with the exception of the Mental Health Threat *S*s who manifested an increase in prejudice toward Negroes but did not show an increase on the general ethnocentrism items.

Another possible consequence of the linkage of the atomic threat to war with the Soviet Union is that some subjects may have responded with hostility to the Russians which they then generalized to other groups. Thus, there is experimental evidence (2) that, under certain conditions, the evocation of aggressive responses to one class of stimuli increases the probability of aggressive responses being made to other classes of stimuli. If this effect were operating in the present instance, one would have to modify the Shared Threat hypothesis to take into account possible differences between shared threats such as atomic war which may call for aggressive reactions, and shared threats such as flood and hurricane which call for more constructive group problem solving activity. Sherif (10) has provided evidence that the latter type of activity helps reduce aggressive tension between groups.

A question is then raised as to what were the essential characteristics of the Flood and Hurricane shared threat that brought about the reduction in prejudice. The initial theoretical discussion stressed the common impact of the shared threat, the fact that the majority and minority group members are equally threatened by impending disaster. However, the critical variable may be the recognition of the need for shared, task-oriented behavior by both Negro and white in order to cope satisfactorily with the danger. It should be feasible to obtain evidence relevant to this issue by employing an experimental design in which the probable value of cooperative group activity with regard to reducing the threat is varied.

The results of the present study and those of the first experiment point up the importance of the interrelationship of motivational and cognitive factors in the determination of prejudiced behavior. They are consistent with the view that social prejudice is not a fixed trait or attitude that only awaits the proper conditions to reveal itself in its purest form. Rather, social prejudice is a consequence of contemporaneous cognitive and motivational factors and varies systematically as a function of variations in these determinants. The present data indicate that variations in the degree and type of threat are among the determinants of prejudiced behavior.

[4] While in terms of the present analysis the cancer threat would appear to be a personal threat, the impersonal manner of its arousal in the original study renders ambiguous its shared-personal threat status. The questions employed to elicit anxiety in regard to cancer were as follows: 1. What would you say is the most serious disease affecting American adults and for which an adequate cure has not yet been found? 2. What are the symptoms of cancer? How would a person know if he has cancer? 3. What are the chances of recovery from cancers such as lung cancer and stomach cancer?

SUMMARY

The present experiment was designed to determine the effects of varying types of threatening communications upon the expression of social prejudice. Specifically, it was hypothesized that a personal threat results in an increase in prejudice while a shared threat results in a decrement in social prejudice. A personal threat was defined as one that is likely to be perceived by a subject as primarily directed toward himself while a shared threat is likely to be perceived as equally affecting members of majority and minority groups.

College students were assigned at random to either the Control treatment, to one of three Personal Threat-Arousal treatments or to one of two Shared Threat-Arousal treatments. A social prejudice questionnaire was administered to the entire sample. Four weeks later the experimental treatments were administered. The experimental variable was manipulated by varying the content of written communications presented just prior to readministration of the social prejudice questionnaires.

The main results were:

1. Each of the three Personal Threat groups showed a greater increase in social prejudice than did the Control group, the difference being significant at the .05 level for two of the groups.

2. Of the two Shared Threat groups, the

Flood and Hurricane Threat group showed a reliable decrease in prejudice in comparison to the Control group while the difference between the Atomic War Threat group and the Control group was not significant.

Since five of the six experimental predictions were confirmed by the data, the results were interpreted as being consistent with the theoretical distinction made between personal and shared threats. The possible influence of other variables on the response to personal and shared threats was discussed and the shared threat variable was further analyzed into a number of possible components.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. FESHBACH, S. The catharsis hypothesis and some consequences of interaction with aggressive and neutral play objects. *J. Pers.* 1956, **24**, 449–462.
3. JANIS, I. L., & FESHBACH, S., Effects of fear-arousing communications. *J. abnorm. soc. Psychol.*, 1953, **48**, 78–92.
4. JANIS, I. L., & FESHBACH, S. Personality differences associated with responsiveness to fear-arousing communications. *J. Pers.*, 1954, **23**, 154–166.
5. MASLOW, A. H. Deprivation, threat and frustration. *Psychol. Rev.*, 1941, **48**, 364–366.
6. MASLOW, A. H., & MITTELMAN, B. *Principles of abnormal psychology*. New York: Harper, 1951.
7. MILLER, N. E., & BUGELSKI, R. Minor studies of aggression: II. The influence of frustrations imposed by the in-group on attitudes expressed toward out-groups. *J. Psychol.*, 1948, **25**, 437–442.
8. MOLTZ, H., & THISTLETHWAITE, D. L. Attitude modification and anxiety reduction. *J. abnorm. soc. Psychol.*, 1955, **50**, 231–237.
9. SAUL, L. J. *The hostile mind*. New York: Random House, 1956.
10. SHERIF, M., HARVEY, O. J., WHITE, B. J., HOOD, W. R., & SHERIF, C. W. Experimental study of positive and negative inter-group attitudes between experimentally produced groups. Robbers Cave study. Norman: University of Oklahoma, 1954 (multilithed).

Received January 24, 1957. Prior publication.

CASE REPORT

THE WOLF

R. K. GREENBANK[1]

Springfield, Delaware County, Pennsylvania

The existence of fantasy playmates in young children is well known to child psychiatrists. Psychotics are also known to suffer from vivid hallucinations. This paper describes an adult nonpsychotic who continued the existence of a fantasied wolf from her childhood.

The Patient

Miss L. was a twenty-four-year-old, single, white, Protestant, professional hospital worker who first sought psychiatric help for severe anxiety and phobic attacks. Face-to-face psychoanalytically oriented psychotherapy on a three-hours-a-week basis was used. Her Rorschach was interpreted as showing,

> a fairly typical record of an hysteric. She shows a great deal of anxiety with a need to escape from her present environment. She is extremely immature emotionally, not realistic or logical. She feels herself appropriating unduly the time of others. Her personal relationships are extremely poor and her friends limited in number.

The patient's intellectual level was high average to superior.

"My Animals"

Early in treatment Miss L. spoke frequently of animals and with more affect than she typically used about humans. "In general, I like animals better than people. If people behaved like animals, this city wouldn't be such a madhouse. People are little, mean, and nasty; dogs aren't." "I feel like an animal on a leash." By her eleventh visit she could say of the therapist, "You watch me like I watch an animal. This is good because you understand me and are not excited by me."

She spent much time discussing pet dogs (real ones) she had owned in childhood. She was much involved emotionally in their deaths. She was particularly concerned with the clipping of one dog's tail. "I'm always close to my dogs. If I don't have one, I feel afraid."

By the 80th hour, she felt therapy was going deep and once said, "You better know what you are doing, for if you take something away from me which I have, I'll die. I just won't be anymore."

After 100 hours of therapy the patient felt sufficient confidence to admit the existence of her fantasied wolf. She had kept this secret from her family, friends, and former therapist. She had previously feared the reaction of people if she discussed this aspect of her life. In a very guarded way over a period of several hours she began to reveal a particularly vivid and affectively colored fantasy life. She first mentioned her fantasy animals of childhood, then the present existence of "My horse and the wolf."

The patient had a large group of fantasied animals in childhood. These included a lion, a bear, a fox, squirrels, dogs, rabbits, etc., all of which were her friends. "They did not like people, but they liked me 'cause I'm different from other people." She lost all of these as she matured.

The two animals that stayed beyond her childhood were "the Wolf" and a horse. The horse can be mentioned briefly, since it never took an active part in the therapy. It pulled automobiles that she rode in; it ran beside trains on the way home. She often imagined riding on him. He was a big white horse which always wore a harness "because I like them." She did not bring him into the city because "horses should run free in the country."

Development of the Wolf

As a child the patient played a game of "Big Black Mouse" with her younger brother. They placed all their possessions on a bed, then the fantasied big black mouse would appear, and they would run in terror to the bed where they were safe. "My brother did not enjoy this much because he did not know what was going on." There was much pleasure in this game but some tinge of fear.

One day while she was playing alone, the mouse got between her and the bed and drove her into a corner; it was going to bite her. The wolf then made its first appearance. It ran in

[1] Student, Philadelphia Psychoanalytic Institute.

the door and gobbled up the mouse and served as her protector thenceforth. She kept the wolf "because I like him and he liked me." "He's an old friend. I think he is as old as I am."

The wolf is "two and a half feet tall, not a pretty color, kinda grey, real nasty to people." He looks almost like a "laughing wolf" but "has great big teeth." "He doesn't wear a harness because he is wild." "He is big, real big. He used to be over my head. He stayed the same size but I grew." "He has lots of fur —too much for a wolf, really."

The wolf was partially under the patient's conscious control, i.e., she could make him come when she especially wanted him, but he often appeared without her asking for him. He often surprised her. "He feels the way I do. If I like a person, he does too; if I don't, he doesn't."

The wolf on at least two occasions was left with other people to care for them. Once she left him with her general physician (unknown to him) and once with her psychiatrist. She had noted the concern of her psychiatrist about a personal problem of his own. She made no comment at the time, but a week later when he felt better she casually asked, "Do you mind if I take my wolf with me? You don't seem to need him anymore." Further investigation revealed that she had felt that "You needed the wolf more than I did." Accordingly, she had left him to take care of the therapist.

In the latter part of her therapy, Miss L. had a real pet dog, Toby, which she had to leave in her apartment. This bothered her for she felt he would be lonely, so she sometimes left the wolf to watch the dog.

The wolf was a major source of protection to the patient. When it was necessary for her to have a major operation, she described the wolf following the stretcher as she went to the operating room. "He was snarling and vicious. He would have torn the doctors to pieces if they did anything wrong." She felt calm but said, "If it were not for him I'd have been scared to death."

The wolf had the same emotions as the patient. She often expressed her feeling through him. "My wolf is so sad since my boy friend went on vacation."

The wolf was a good indicator of the state of transference. At the first interview, the wolf sat firmly fixed between the patient and the therapist. He was "alert, watching every move you made." Later he relaxed and often slept through the interview at the patient's feet. If the patient was under emotional strain or charged material was being approached, the wolf would raise his head and growl.

About a previous therapist, whom the patient had strongly disliked because of his aggressive manner, she said, "My wolf never rested when I was with Dr. M. He ran around the office, snarling, howling, and growling and his eyes were so red."

The wolf kept most frequently on the right side of the patient. He watched closely to see what the patient wanted. "He's glad to be there."

Early in therapy there was an occasion when the therapist found himself in close physical proximity to the patient on a train. He noted a sense of anxiety and concern in the patient and believed this was a result of the transference aspects of therapy and the physical closeness. He moved away and she showed signs of considerable relief. She never discussed this problem, although several openings were made by the therapist. After she had described the wolf to the therapist six months later, she then admitted that he had caused her a great sense of anguish, since he was standing on the wolf's tail. The fantasy animal was howling, but she felt unable to ask the doctor to move.

As the patient proceeded along well in her therapy, the wolf came less and less into notice. She fell in love with the president of a fraternity in a medical school in the same city. As this relationship developed, the wolf disappeared gradually from her life.

Earlier she had said, "It's nice to have you (the therapist) and the wolf, but I want a real person who talks to me, thinks about me, is nice to me, and whom I don't have to imagine." She found this in her boy friend.

After the wolf left the last time, she said, "He was chained up long enough to somebody who was unhappy and scared, and now he wants to run loose for awhile." But she missed him. "I do not know where my poor wolf is; I haven't had him for a long time. He must be around, but I do not know where."

DYNAMIC CONSIDERATIONS

Her fantasy animals, especially the wolf, probably prevented the development of a frank psychosis and allowed Miss L. to func-

tion on a neurotic level. This was, however, a fluid state. She developed paralysis, blindness, globus hystericus, tunnel vision, and *la belle indifference* under moderate stress. If the strain became severe, especially with emotional deprivation, she regressed temporarily to more malignant symptoms such as withdrawal, paranoid thoughts, adult enuresis, and delusional thinking.

There was considerable emotional deprivation in her early life. Her neurotic mother and alcoholic father were unable to give her adequate love. Her loving and devoted grandfather died suddenly in the patient's presence. As a child, she was severely punished and rejected when she showed normal aggressive feelings.

By using the wolf as an auxiliary ego, she was able to retain the essential aggression without suffering the guilt of being aggressive herself. Both the horse and the wolf, being symbols of masculine strength and power, provided her with a substitute for her weak father.

She, therefore, turned away from the emotional deprivation of her early life in the autistic pleasures to be gained in fantasy with her animals.

SUMMARY

The case of a young woman has been presented who continued into adult life a hallucinated wolf from her childhood. It remained her support and companion until she was able, through therapy, to develop a heterosexual relationship with an accepting male.

Received July 17, 1956.

CRITIQUE AND NOTES

A GUTTMAN SCALE FOR SEXUAL EXPERIENCE—A METHODOLOGICAL NOTE[1]

LAWRENCE PODELL AND JOHN C. PERKINS

The City College of New York

Veterans Administration Mental Hygiene Clinic, Buffalo, N. Y.

In some investigations of sexual behavior, there is an implication that heterosexual experience cumulates in a certain sequential pattern. For example, Bromley and Britten (2, p. 290) employed the following item: "If you care for hot petting, how far do you go?" Kinsey, Pomeroy, and Martin (3) state that:

> Petting techniques usually expand in a more or less standard sequence, as the partners become better acquainted. Beginning with general body contact, lip kissing, and the deep kiss, it advances to a deliberate manipulation of the female breast, to mouth contacts with the female breast, to manual stimulation of the male genitalia, and finally to oral stimulation of the female genitalia (3, p. 540).

This statement implies a cumulative sequence wherein a person who has participated in a "more advanced" behavior may be assumed to have experienced the "less advanced" behaviors. In other words, the statement suggests that a male who has had oral contact with the female breast has also engaged in lip kissing, deep (tongue contact) kissing, and manual stimulation of the female breast.

Does heterosexual experience in fact cumulate along a unidimensional continuum? If so, what is the sequence of behaviors? The present study is directed to these questions. To the knowledge of the authors, such a unidimensional cumulative scale of sexual behaviors has not been established in any studies thus far published.

Method

The responses used in this research derived from personal interviews. One hundred unmarried male college students, 50 with extremely unfavorable marital prediction scores and 50 with extremely favorable marital prediction scores on the Adams' Marital Happiness Prediction Inventory (1), volunteered to participate. The subjects were between 18 and 22 years of age. Fifty-eight reported themselves as Protestants, 24 as Roman Catholics, 15 as Jewish, and the remainder claimed no religious affiliation. Fifty-eight were raised in a rural environment, 35 in an urban environment, and the remainder in mixed settings.

Results

The Cornell technique of Guttman Scale Analysis (4) was used to analyse the *Ss*' responses to 16 items concerning heterosexual behavior.[2] Of these items 15 were found to form a Guttman scale.[3] The derived order of items appears in Table 1, with the number of errors and nonerrors presented for each item.

To determine the degree to which the derived scale deviated from a theoretically perfect Guttman scale, the coefficient of reproducibility was computed.[4] Technically, the proper procedure in Guttman scaling calls for the elimination of items involving frequencies of 10 per cent or less and 90 per cent or more. Four of the 15 items (items 1, 2, 3, and 15 in Table 1) were in these categories. Although these items were used in Table 1 for illustrative purposes, they were not included in computing the coefficient of reproducibility for the scale. The coefficient of reproducibility is .935. Thus, knowing the subjects' scores, we can state those behaviors which the subjects have reportedly experienced with over 93 per cent accuracy. (The coefficient of reproducibility for all 15 items is .948.) The distribution of responses forms the

[1] The data used in this paper were obtained in the course of a doctoral research by J. C. Perkins (5).

[2] The sixteen items were selected from a total of 271 items covering a wide variety of sexual attitudes and behavior. Questions 88 to 93 in the interview concerned (in the order given) close embrace, lip kiss, tongue kiss, fondling the clad breast, fondling the nude breast, kissing the breast, sucking the breast, caressing the female sex organs, fondling the male nipples, caressing each other's sex organs, fondling of the male sex organ to orgasm, bare genital contact without intromission, and interfemoral relations. Two items followed concerning the effects of petting. Question 96 concerned sexual intercourse. Fifty-eight items followed, after which the questions concerning oral contact with the male genitalia and female genitalia were asked. The actual wording of the questions may be found in the original study (5, pp. 253-266).

[3] It was found that item 16, "In a necking or petting situation, has a girl ever played with your nipples" could be substituted for item 15, "oral contact with the female genitalia," and still fulfill the requirements of a Guttman scale. However, if both items are included in the attempt to derive a scale, either one, depending upon the ordering of respondents, will incur more errors than nonerrors and must be eliminated. In other words, either behavior is part of a unidimensional sequence, but not both. It is possible that different dimensions are introduced at this extreme point, but there were insufficient data to test this possibility.

[4] The coefficient of reproducibility $= 1 - \frac{\text{number of errors}}{\text{number of items} \times \text{number of respondents}}$.

TABLE 1
A GUTTMAN SCALE OF PREMARITAL HETEROSEXUAL EXPERIENCE FOR 100 COLLEGE MALES

Sequence	Type of Behavior Reported	Errors	Non-errors
15	Oral contacts with female genitalia	6	10
14	Oral contacts with the male genitalia	7	21
13	Interfemoral relations	7	34
12	Bare genital contact, without intromission	6	37
11	Sucking of the female breast	15	39
10	Manual manipulation of the male genitalia to orgasm, by the female	10	51
9	Sexual intercourse	8	49
8	Manual manipulation of each other's genitalia	3	63
7	Manual manipulation of the female genitalia, by the male	2	68
6	Kissing of the female breaat	9	63
5	Manual manipulation of the nude female breast, by the male	1	73
4	Kissing with tongue contact	3	84
3	Manual manipulation of the clad female breast, by the male	1	91
2	Lip kiss	0	99
1	Embrace	0	99

parallelogram pattern characteristic of the Guttman scale.

DISCUSSION

In published researches, there seems to be a tendency to consider sexual intercourse as the most extreme behavior in a sequence of sexual behaviors. For example, the phrases "complete sex experience" and "go the limit" are apparently used as synonymous with sexual intercourse (7, pp. 61 and 247). The arbitrary judgment of any particular behavior as "most extreme" by an investigator may indicate personal biases on his or her part. This research, by objectively and empirically ordering premarital heterosexual experience along a unidimensional scale, would seem to suggest that sexual intercourse is not the most extreme type of sexual behavior, though it tends to be more extreme than some other types.

Not only do existing studies imply a cumulative sequence of sexual experience, but they also imply that persons experience the "less extreme" behaviors before the "more extreme" ones. This is an assumption which we cannot make; although a temporal pattern of experience will be cumulative, the existence of a Guttman scale does not prove a temporal sequence, e.g., the respondents who experienced item 12 may have done so before or after item 9. To demonstrate a temporal sequence, the questions would have to be phrased accordingly.

A meaningful interpretation of the sequence of behaviors appearing in Table 1 awaits further research. That a single dimension of sexual experience was discovered rather than a plural number of dimensions (e.g., an oral dimension, a genital dimension) remains to be explained. It is possible that the ordering of the reported behavior is influenced by the factor of the "social acceptability" of the behavior. For example, the discrepancy in rank order between items 6 and 11 may be a matter of the respondent's perception of sucking the breast as an "infantile" behavior. (Another possible explanation for this rather outstanding discrepancy may be offered: The question concerning item 6, "kissing the female breast," did not specify the nipple of the breast and it was not ascertained whether or not the respondents interpreted the question to include the nipple. The same holds true for item 11, "sucking the female breast." However, it seems more probable that nipple contact was greater in the behavior of item 11 than in that of item 6; the difference in rank order of the items in the scale.)

It should be noted that *S*s in this study were a highly selected sample. Therefore, the scale presented in this note may not be applicable to single men in general, nor even to single men whose backgrounds are similar to the backgrounds of the respondents used in this study. Nevertheless, the fact that a scale could be obtained from two extreme groups seems to suggest that it is feasible to develop satisfactory scales of sexual behavior, which may or may not be similar to the one presented here. Comparison of the incidences of sexual behaviors of this highly selected sample with the findings of Kinsey, Martin and Pomeroy (3) for a sample of corresponding age and educational level revealed little difference. The comparison may be found in Perkins' dissertation (5, pp. 267–268). Using other populations, we might expect to find different scale patterns or even the absence of unidimensionality. In the light of the studies by Kinsey et al (3) it would appear that in using populations which differed from the sample in socioeconomic status, educational level, or age range, we might expect differences in sexual experience (e.g., a differential distribution of petting behaviors by educational level) and the scalability of such experience.

In addition, it must be noted that the Guttman scale illustrated in this research need not be identical with a scale derived from the responses of females, even respondents with similar backgrounds. In fact, there is evidence available to suggest that this is not the case (6).

SUMMARY

The reported premarital heterosexual sexual experience of 100 highly selected college men was subjected to the Cornell Technique of Guttman

Scale Analysis, with a satisfactory scale forthcoming. It therefore appears feasible to order certain aspects of sexual behavior along a unidimensional cumulative scale, thus aiding in the definition of premarital sexual experience.

REFERENCES

1. ADAMS, C. R. *The marital happiness prediction inventory.* State College: Pa. State College, 1950.
2. BROMLEY, D. D., & BRITTEN, F. *Youth and sex.* New York: Harper, 1938.
3. KINSEY, A. C., POMEROY, W. B., & MARTIN, C. E. *Sexual behavior in the human male.* Phila.: W. B. Saunders, 1948.
4. GUTTMAN, L. The Cornell technique of scale and intensity analysis. *Educ. psychol. Measmt,* 1947, 7.
5. PERKINS, J. C. Marital prediction scores of college men relative to behavior and attitudes. Unpublished doctor's dissertation, Pennsylvania State Univer., 1954.
6. REEVY, W. R. The differential responses of women with a favorable marital happiness prediction and women with an unfavorable marital happiness prediction. Unpublished doctor's dissertation, Pennsylvania State Univer., 1954.
7. ROCKWOOD, L. T., & FORD, M. E. N. *Youth, marriage and parenthood.* New York: J. Wiley, 1945.

Received June 18, 1956.

LEVELS OF PERCEPTUAL ORGANIZATION AND COGNITION: CONFLICTING EVIDENCE[1]

E. W. PICKREL

Personnel Laboratory, Air Force Personnel and Training Research Center

RATHER low relations have generally been found between performance on perceptual and cognitive tasks (4, 5). However, Krech and Calvin (3) have recently reported a surprisingly high relation, a biserial correlation of .91 with a standard error of .07, between a culture-free perceptual variable and the culture-laden intelligence variable. Their perceptual variable was the extent to which a dotted pattern, presented tachistoscopically at .06 second, had been articulated; that is to say, whether a subject organized the stimulus (see Fig. 1) into a homogeneous pattern of equidistant dots or whether he left at least twice as much space between the rows as within the rows (or columns) of dots. As a measure of intelligence they selected the vocabulary test from the Wechsler Bellevue Intelligence Scale. It was suggested that this high relation was a demonstration of continuity between perceptual and cognitive processes, but the magnitude of their obtained correlation is at variance with that commonly found between such tasks. The purpose of this study is to reproduce their experiment to verify the results and observe the relation of this new measure to those perceptual variables most commonly used in studies of individual differences.

METHOD

Apparatus and Measures

The apparatus included a tachistoscope, stimulus cards, a "reproduction board" and a supply of small black plastic beads. The reproduction board, plastic beads, and their use have been previously (3) described. The stimulus cards were presented by means of an electronically controlled Dodge-type mirror tachistoscope. The stimulus fields were illuminated by fluorescent lamps to a level of .61 foot candle and the illumination of the pre- and post-exposure field was matched with exposure field at the point of the observer's eye. The stimulus cards (see Fig. 1) were photographic copies from an original used by Dr. Calvin. The sole criterion of a successful reproduction[2] was that for any single trial there should be at least twice as much space between the rows as within the rows (or columns) of the patterns.

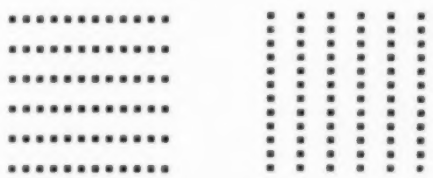

FIG. 1. STIMULUS CARD

Following the tachistoscopic measure subject (*S*) was administered three of the Keystone Visual Skills tests, the vocabulary test taken from Wechsler (7), and three paper-and-pencil tests. The Keystone tests were for determining whether there was a significant relation between the perceptual phenomenon and visual skills, the paper-and-pencil tests for determining if articula-

[1] This report is based on work done under ARDC Project No. 7719, Task No. 17011, in support of the Research and Development Program of the Air Force Personnel and Training Research Center, Lackland Air Force Base, Texas. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government.

[2] As stated in a personal communication by Dr. Calvin.

tion was related to the three commonly used factor measures of individual differences in perceptual abilities. The tests selected from the Keystone battery are "near point" measures for lateral imbalance, fusion, and usable vision. The three paper-and-pencil tests are measures of the principal spatial ability factors that have been isolated (1). Pattern Comprehension Test BP803C yields a measurement of how quickly *S* can understand how the pattern of an object will fold into a solid figure. Rotated Figures Test BP112C is a measure of how quickly *S* can identify a stimulus card when it is rotated in two-dimensional space. Fish Bowl Test UT-19FB is a measure of the ability to determine rapidly spatial relations relative to the body orientation of the observer. The *S* is given a head-on view of an object floating in a fish bowl, then given a different view of the same object and must determine the new position to which he must project himself to obtain this second view of the object.

Subjects

To approximate the ability level of the sample used by Krech and Calvin, airmen with percentile scores of 85 and above on the Armed Forces Qualifying Test, a general intelligence measure, and AFROTC graduates attending the Officers Military School at Lackland Air Force Base were selected as *S*s. One hundred and twenty-four *S*s were used for the experiment.

Procedure

In a pilot study which used the directions and exposure time of the original experiment, no significant relation was found between performance on the perceptual and verbal measures. Similar results have been obtained concurrently by De Soto and Leibowitz (2). It was found from protocols that *S*s had been seeking extraneous cues on which to focus their attention, and were not considering a solution as inherent in the stimuli being observed tachistoscopically. For this reason the instructions were rewritten to focus attention on the stimulus cards and accurate reproduction of perceptions.

Each *S* was given the following instructions at the beginning of his first experimental session:

> I am going to present you with a test of visual acuity, that is how well you can observe. You must realize that visual acuity is becoming more and more important as our aircraft increase in speed. We are asking that you serve as guinea pigs, as pilots, for us while we are developing this measure. You will be presented with two symbols for a very brief exposure on the screen. For any given trial these symbols may seem to be alike or they may appear to be different. They may seem to change from trial to trial, or they may appear to be the same. After each presentation I want you to reproduce both of the symbols as accurately as you can by simply drawing them on this board with these beads as this would be a drawing of the five on dice (demonstrate). Now look into the box. You see a cross which is roughly in the center of a white field. One of the symbols will be on the left-hand side of the cross, the other on the right. They will be somewhat smaller than you can draw on the board, but make them proportional.
>
> To repeat the directions, you will be presented with two symbols for a very brief exposure on the screen. The time is so short that I want you to say "ready" each time. Then the symbols will appear simultaneously. Remember that for any given trial these symbols may seem to be alike or they may appear to be different. They may seem to change from trial to trial, or may appear to be the same. After each observation draw both symbols as accuratedly as you can on the board and I will tell you whether your drawing is correct. Are there any questions?

Questions were answered in the same manner as in the original experiment. After each trial *S* was told if his percept was correct or incorrect for the purpose of directing closer and closer attention to stimulus detail. The procedure was continued until 16 trials had been completed.

TABLE 1

VISUAL SKILLS TEST PERFORMANCE FOR SUBJECTS IN EACH GROUP

Levels of Perceptual Organization	Fusion			Convergence			Visual Acuity	
	N	Normal	Deviate	Over	Normal	Under	20/40 to 20/22	20/20 to 20/15
Nonarticulated Group	74	73	1	11	60	3	8	66
Articulated Group	50	49	1	7	39	4	6	44

TABLE 2

INTERRELATION OF THE PERCEPTUAL AND VERBAL MEASURES

	Paper-and-Pencil Measures			Level of Perceptual Organization
	Pattern Comprehension	Rotated Figures	Fish Bowl	
Vocabulary Test	.14	.03	.22*	.34**
Level of Perceptual Organization	.15	.10	.12	

* Significant at the .05 level.

** Significant at the .01 level.

Note.—Vocabulary—paper-and-pencil coefficients are product-moment correlations, the other four coefficients are biserial correlations.

RESULTS

It is evident from Table 1 that there was no significant relation between the control measures of visual skills and levels of perceptual organization. The measure of visual acuity yielded an $F_{1/2}$ of .05 with the articulated-nonarticulated pattern groups.

The interrelations of perceptual and verbal measures are presented in Table 2. The correlation between vocabulary scores and dichotomized levels

of perceptual organization was .34 with a σr_{bis} of .10. Using a test presented by Walker and Lev (6) it was found that the difference in magnitude between the .34 coefficient reported here and the .91 coefficient obtained by Krech and Calvin was significant at the .01 level.

Mean score on the vocabulary test for the 50 *Ss* who reached an articulated level was 28.64 with an *SD* of 3.76. Twenty of these 50 *Ss* had scores of 30.5 or higher on the vocabulary test, a cutting point which differentiated the articulated and nonarticulated groups in the original experiment. Mean score for the 74 *Ss* holding to the homogeneous percept was 26.16 with an *SD* of 4.59. Fifteen of these *Ss* had scores of 30.5 or higher on the vocabulary test.

SUMMARY

Though rather low relations have generally been found between performance on perceptual and cognitive tasks, Krech & Calvin recently reported a surprisingly high relation between perception of a dotted pattern viewed tachistoscopically and performance on the Wechsler vocabulary test. The purpose of this study was to reproduce their experiment to verify the results and observe the interrelation of the measures employed to those perceptual variables most commonly used in studies of individual differences. A slightly negative relation was found between performance on the perceptual and cognitive tasks for an initial sample. Since an artificial problem contained within the original administrative directions seemed to distract *Ss* from the perceptual task, administrative directions were rewritten to focus each *S*'s attention onto the perceptual problem. The Krech measure yielded a higher relation than the other perceptual measures to verbal test performance and had a very low relation to the other perceptual tests.

The size of the correlation reported here between the Krech measure and verbal test performance was significantly lower than that obtained by Krech & Calvin in an earlier study. However, the correlations obtained in this study between vocabulary scores and perceptual test performance are of a size often found between perceptual and verbal measures in studies of individual differences and are more consistent with such information than that found by Krech and Calvin.

REFERENCES

1. ANDERSON, G. V., FRUCHTER, B., MANUEL, H. T., & WORCHEL, P. Survey of research on spatial factors. *Research Bull. AFPTRC-TR-54-84*, 1954, AF Personnel & Train. Res. Cent., Lackland Air Force Base, San Antonio, Texas.
2. DE SOTO, C., & LEIBOWITZ, H. Perceptual organization and intelligence: A further study. *J. abnorm. soc. Psychol.*, 1956, **53**, 334–337.
3. KRECH, D., & CALVIN, A. Levels of perceptual organization and cognition. *J. abnorm soc. Psychol.*, 1953, **48**, 394–400.
4. PICKREL, E. W. Prediction of the trainability of slow learners from tests with a symbolic and nonsymbolic content. *Res. Bull. AFPTRC-TR-54-82*, 1954, AF Personnel & Train. Res. Cent., Lackland Air Force Base, San Antonio, Texas.
5. THURSTONE, L. L., & THURSTONE, THELMA G. Factorial studies of intelligence. *Psychometr. Monogr.*, 1941, No. 2.
6. WALKER, HELEN, & LEV, J. *Statistical Inference*. New York: Henry Holt & Co., 1953.
7. WECHSLER, D. *The measurement of adult intelligence*. (2nd ed.) Baltimore: Williams and Wilkins, 1941.

Received June 25, 1956.

AUTHORITARIANISM OR ACQUIESCENCE IN BASS'S DATA[1]

SAMUEL MESSICK AND DOUGLAS N. JACKSON[2]

Educational Testing Service and Princeton University — *Pennsylvania State University*

SOME errors have been located in a recent article in this journal (1) which necessitate a reinterpretation of the results. In his study of authoritarianism and acquiescence, Bass obtained a correlation of −.20 between the California F scale and an F scale rewritten so that the content of each item reflected a viewpoint opposed to the original (called the G scale). Bass used this correlation in a factor analysis of the F and G scales, placing reliabilities in the diagonals. The starting point for the analysis was a correlation matrix between F and −G [Table 4 in (1)]. In this matrix the G scale had been reflected and $r_{FG(-)} = .20$ was used instead of $r_{FG} = -.20$. However, the notation for the factor matrix of Table 5 (1) was in terms of F and G, not F and −G. Since the

[1] Most of these remarks were made in private correspondence with Bernard M. Bass. They are presented as a separate article at the suggestion of the editor in the interest of clear presentation of the corrections and interpretations.

[2] The authors wish to thank Dr. Milton Rokeach for first pointing out the disparity between Bass's data and his conclusions and Dr. Wayne H. Holtzman for participating in our joint critical analysis of Bass's article.

factor loadings in Table 5 reproduce a correlation of $r_{FG} = +.20$, the factor matrix labels should read F and $-$G just like the correlation matrix labels. Thus, F and $-$G, not F and G, have positive loadings on the first factor, necessitating a reversal in interpretation.

This outcome may also be seen by factoring the following unreflected correlation matrix with reliabilities in the diagonals:

	F	G
F	(.81)	$-.20$
G	$-.20$	(.50)

The factor matrix which reproduces these correlations is

	I Authoritarianism	II Acquiescence
F	.77	.47
G	$-.53$	.47

and not Bass's Table 5, which is identical except for reversing signs in the second row and reversing the labels of the factors. Now, approximately three-fourths of the reliable variance of the F scale is apparently associated with an *authoritarianism* factor while only one-fourth is attributed to *acquiescence*.

Since these factors are based upon a single correlation coefficient, they are certainly not determined in the usual factor analytic sense. Ordinarily, at least three tests and preferably more are necessary to define a single common factor in Thurstone's system and at least five tests are required to determine two common factors (5). Also, since Bass used reliabilities instead of communalities in the diagonals, these factors involve both specific and common variance, thus making the interpretation particularly difficult. The label "acquiescence" attached to the second factor is a very reasonable interpretation, but there is no way of ascertaining whether Factor II might not be a second content dimension due to an inadequate reversal of the scales or perhaps even some confounding of specific content factors and acquiescence.

On the other hand, it should be pointed out that Bass was probably not attempting to isolate and establish psychologically meaningful factors; he was using factor analysis to *describe* the reliable variance of a scale and to show that one factor was not an adequate description. Another way of demonstrating this would be to use the reliabilities to correct for attenuation. If the unattenuated correlation between the two scales is significantly less than -1.00 (as it is in this case), then more than one factor is represented. Again, there is no way of determining the nature of the second factor from this single correlation coefficient. Although acquiescence provides the most likely hypothesis, the possible operation of additional content factors cannot be discounted.

Bass's interpretation of the analysis of variance presented in his article suggests that the variance attributable to acquiescence may still be significant. However, this analysis is based upon the assumption that if no response set existed, each individual's F scale score when combined with his G scale score should sum to zero. This would be true if the G scale items were perfect reversals in direction and equally potent in opposition to the corresponding F scale items. However, the various levels of opposition reflect imperfect reversals, and the different variances for the F and G scales indicate that they were not opposing parallel forms of the kind necessary to cancel scores. Since the basic assumption was not met, the analysis of variance is difficult to interpret.

There is also the possibility that agreements to corresponding original and reversed F scale items might indicate that the statements are not really opposite in meaning. The correlations between F and G range from $-.52$ to $+.05$ going from the best level of opposition to the worst. The r_{FG} of $-.52$ is in the direction predicted by consistent responses to content in the two scales, and it approaches the theoretical maximum correlation that would be expected given the present reliabilities. As the judged adequacy of the reversals decreases, r_{FG} becomes more and more positive, again as would be predicted in terms of content responses to items whose meanings are becoming less perfectly opposed. However, it should also be pointed out that the high negative correlation is associated with items reversed by inserting "not" or "no" into the original statement and that opposing responses would be expected to such item pairs. Acquiescence, on the other hand, has a much better chance of appearing in those items which have similar styles of expression yet are obviously opposed in meaning without being simple reversals. Such items might be clearly opposite semantically and still be judged less opposite than a simple reversal.

The point to be emphasized is that the over-all correlation of $-.20$ in this sample is in the content direction, but it is not as high as a complete content interpretation would require. Bass attempted to demonstrate using analysis of variance and factor analysis that more than a simple authoritarian-content dimension was necessary to account for this intercorrelation in light of the obtained reliabilities. An error in factoring resulted in too much variance being attributed to "acquiescence," but this error does not eliminate acquiescent response set as a possible determinant of F scale scores.

Even though Bass's conclusions do not necessarily follow from his data, his emphasis upon acquiescent response set in the F scale is important and is supported by other evidence. In some recent

studies, correlations between agreement to original F scale items and to various reversed forms were not found to be high and negative (4) as would be expected from consistent responses to item content and were sometimes found to be significantly positive (2, 3). Thus the operation of a consistent response set in the California F Scale has been repeatedly demonstrated and perhaps, as the latter two studies suggest, even a major portion of the variance may be legitimately attributed to acquiescence.

REFERENCES

1. BASS, B. M. Authoritarianism or acquiescence? *J. abnorm. soc. Psychol.*, 1955, **51**, 616–623.

2. JACKSON, D. N. & MESSICK, S. J. A note on "ethnocentrism" and acquiescent response sets. *J. abnorm. soc. Psychol.*, 1957, **54**, 132–134.

3. JACKSON, D. N., MESSICK, S. J., & SOLLEY, C. M. How "rigid" is the "authoritarian?" *J. abnorm. soc. Psychol.*, 1957, **54**, 137–140.

4. LEAVITT, H. J., HAX, H., & ROCHE, J. H. "Authoritarianism" and agreement with things authoritative. *J. Psychol.*, 1955, **40**, 215–221.

5. THURSTONE, L. L. *Multiple-factor analysis*. Chicago: Univer. Chicago Press, 1947.

Received December 4, 1956. Prior publication

REPLY TO MESSICK AND JACKSON'S COMMENTS ON AUTHORITARIANISM OR ACQUIESCENCE

BERNARD M. BASS

Louisiana State University

IN THE light of Messick and Jackson's (3) corrections, for which I am grateful, and subsequent independent studies on the same subject, it appears that the hypothesis advanced in "Authoritarianism or Acquiescence" (1) was more satisfactory than the analysis used to test the hypothesis.

FURTHER COMMENTS ON METHOD

I should like to add the following methodological notes: The statement that any individual's F and G score cancel each other was not correct, but the analysis of variance was another way of showing that F correlated −.20 with G. At the same time, the analysis permitted the examination of the more complex effects of "levels of opposition."

The variance of the F + G scores is given by:

$$\sigma^2_{F+G} = \sigma^2_F + \sigma^2_F - 2r_{FG}\sigma_F\sigma_G$$

The relation or ratio between the F + G variance and the "within-scale" F and G variances is a function of the correlation between F and G. Then the F test, $\sigma^2_{(F+G)}/(\sigma^2_F + \sigma^2_G)$ merely reflects the size and direction of the correlation between F and G.

Two common factors and no specific variances were conceived in the F and G data rather than one common factor and two specific variances, for if F and G covered the same content, with the obtained reliabilities, then they should have correlated −.64 instead of −.20. Either the lack of correlation between F and G was due to specific variances in each, or to acquiescence. The opposition among the items made it difficult to consider the items as measuring different rather than opposite content.

Subjects and Regional Differences

As Messick and Jackson noted, the correlation of −.20 between F and G may have been unusual and by itself would suggest according to the corrected factor analysis that only about 25 per cent of the reliable variance of the F scale can be attributed to acquiescence. In reversed F scale studies carried out with other samples, and by other investigators including Messick and Jackson, correlations ranging from −.42 to +.66 have been found when a high negative relation consistently should have resulted between the F scale and a reversed F scale. Regional factors may be one source of variation in these correlations. Another may be different ways of reversing F scale items. Reversed items may assess different rather than opposing content.

In analyzing the relation between the F scale and a scale of social acquiescence, the author (2) found large regional differences. This work and the varying results of reversed F scale studies suggest that it is impossible to draw quantitative conclusions concerning authoritarianism and acquiescence without specifying the origins of the sample used as well as other conditions. Of three regions examined, Kansas, Maryland, and Louisiana, the lowest correlation between authoritarianism and acquiescence was found in the Louisiana sample (2). The correlations for Kansas, Maryland, and Louisiana were respectively: .48, 49 and .16 The means in F for the three samples

were: 3.92, 4.03, and 4.52. The means in social acquiescence followed the identical order: 28.7, 29.9 and 33.9. There was a correlation in means between groups as well as a correlation of scores within groups between acquiescence and authoritarianism. Since a Louisiana sample was employed in the study in question (1), the results as they now stand corrected may be more conservative than other investigators such as Jackson and Messick have obtained or will be likely to obtain with other samples.

Throw out the bathtub and the bath water, but save the baby.

REFERENCES

1. BASS, B. M. Authoritarianism or acquiescence? *J. abnorm. soc. Psychol.*, 1955, **51**, 616–623.
2. BASS, B. M. Development and evaluation of a scale for measuring social acquiescence. *J. abnorm. soc. Psychol.*, 1956, **53**, 296–299.
3. MESSICK, S. & JACKSON D. Authoritarianism or acquiescence in Bass' data. *J. abnorm. soc. Psychol.*, 1957, **54**, 424–426.
4. Symposium: Redefining authoritarianism. *Amer. Psychol.*, 1956, **11**, 365(a). (Abstract.)

Received November 26, 1956. Prior publication.

ERRATA

In the article "Rorschach and TAT indices of homosexuality in overt homosexuals, neurotics, and normal males" by Anthony Davids, Mark Joelson, and Charles McArthur in the September 1956 issue of this Journal (*J. abnorm. soc. Psychol.*, 1956, **53,** 161–172), there is an error in the text. In the section on TAT results, under the heading of *Signs suggested for further confirmation* (p. 168), it states incorrectly (line 15) that the sign of a *strong unresolved attachment to a father or father figure* was scored in stories composed for Card 8. The sign was in fact scored in stories composed for TAT Card 7.

In the article entitled "Ethnic cleavage among young children" by W. E. Lambert and Y. Taguchi (this Journal 1956, **53,** 380–382) the following table should be substituted for Table 1 (p. 381), and one p value, for Picture-Taking Choices, has been corrected. The tests were one-tailed since there was no apriori ground for expecting out-group preferences.

TABLE 1

TEST CHOICES OF OCCIDENTAL AND ORIENTAL CHILDREN

	Candy-Giving Choices		Picture-Taking Choices		Picture-Choosing Choices	
	Occid.	Orient.	Occid.	Orient.	Occid.	Orient.
Occidentals	4	3	5	2	4	3
Orientals	0	6	0	6	0	6
Exact tests	$p = .048$		$p = .016$		$p = .048$	

The original tests of significance were calculated from the table in the form given above. Confusions of interpretation may arise from a reading of the original Table 1. The original table implies that the test is being made on the null hypothesis that in-group choice is independent of the ethnic background of the choosers. In reality, we were interested in determining whether there was a preponderance of in-group choices, that is, a tendency for persons to choose their own ethnic group.

The Journal of Abnormal and Social Psychology

Editor: M. BREWSTER SMITH, *New York University*

Associate Editor: E. J. SHOBEN, JR., *Teachers College, Columbia University*

CONSULTING EDITORS

GORDON W. ALLPORT
Harvard University

URIE BRONFENBRENNER
Cornell University

DONALD T. CAMPBELL
Northwestern University

LAUNOR F. CARTER
The RAND Corporation

RICHARD CHRISTIE
Columbia University

MORTON DEUTSCH
Bell Telephone Laboratories

I. E. FARBER
State University of Iowa

LEON FESTINGER
Stanford University

DONALD W. FISKE
University of Chicago

ROBERT R. HOLT
New York University

HERBERT H. HYMAN
Columbia University

IRVING L. JANIS
Yale University

HAROLD E. JONES
University of California

HAROLD H. KELLEY
Center for Advanced Study in the Behavioral Sciences

LEONARD S. KOGAN
Community Service Society of New York

GARDNER LINDZEY
Syracuse University

DAVID C. MCCLELLAND
Harvard University

QUINN MCNEMAR
Stanford University

LEO POSTMAN
University of California

NEVITT SANFORD
University of California

SEYMOUR SARASON
Yale University

RICHARD L. SOLOMON
Harvard University

RENATO TAGIURI
Harvard University

Managing Editor: ARTHUR C. HOFFMAN

Assistant Managing Editor: HELEN ORR

Editorial Staff: BARBARA CUMMINGS; FRANCES H. CLARK; SADIE J. DOYLE

VOLUME 54
1957

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.
MT. ROYAL & GUILFORD AVES., BALTIMORE 2, MD.
AND 1333 SIXTEENTH STREET N.W., WASHINGTON 6, D.C.

Copyright 1957 by the American Psychological Association, Inc.

CONTENTS OF VOLUME 54

ARTICLES, CASE REPORTS, CRITIQUE AND NOTES

PAGE

ARGYLE, M. **Social Pressure in Public and Private Situations** 172
AUSTRIN H. See SCODEL, A.
BAGBY, J. W. **A Cross-Cultural Study of Perceptual Predominance in Binocular Rivalry** 331
BAKER, BELA O., & BLOCK, J. **Accuracy of Interpersonal Prediction as a Function of Judge and Object Characteristics** 37
BARRY, H., III. **Relationships Between Child Training and the Pictorial Arts** 380
BASOWITZ, H., & KORCHIN, S. J. **Age Differences in the Perception of Closure** 93
BASOWITZ, H. See KORCHIN, S. J.
BASS, B. M. **Reply to Messick and Jackson's Comments on Authoritarianism or Acquiescence** 426
BEHAR, I., & RIOPELLE, A. J. **The Effects of Reserpine on the Acquisition of the Conditioned Avoidance Response** 160
BELL, R. Q. See SCHAEFER, E. S.
BENSEN, S. B. See SEIDMAN, D.
BENTON, A. L., & BLACKBURN, H. L. **Practice Effects in Reaction-Time Tasks in Brain-Injured Patients** 109
BERKOWITZ, L. **Liking for the Group and the Perceived Merit of the Group's Behavior** 353
BERRY, J. L., & MARTIN, B. **GSR Reactivity as a Function of Anxiety, Instructions, and Sex** 9
BLACKBURN, H. L. See BENTON, A. L.
BLOCH, BARBARA L. See MAUSNER, B.
BLOCK, J. **A Comparison Between Ipsative and Normative Ratings of Personality** 50
BLOCK, J. See BAKER, BELA O.
BLOCK, J. **Studies in the Phenomenology of Emotions** 358
BRACKBILL, YVONNE, & LITTLE, K. B. **Factors Determining the Guessing of Meanings of Foreign Words** 312
BRIM, O. G., JR., & HOFF, D. B. **Individual and Situational Differences in Desire for Certainty** 225
BROEN, W. E., JR. **A Factor-Analytic Study of Religious Attitudes** 176
BROWN, R. W., LEITER, R. A., & HILDUM, D. C. **Metaphors from Music Criticism** 347
BURWEN, L. S., & CAMPBELL, D. T. **The Generality of Attitudes Toward Authority and Nonauthority Figures** 24
CAMPBELL, D. T. See BURWEN, L. S.
CAMPBELL, D. T. See CHAPMAN, L. J.
CATTELL, R. B., STICE, G. F., & KRISTY, N. F. **A First Approximation to Nature-Nurture Ratios for Eleven Primary Personality Factors in Objective Tests** 143
CERVIN, V. **Relationship of Ascendant-Submissive Behavior in Dyadic Groups of Human Subjects to Their Emotional Responsiveness** 241
CHAPMAN, L. J., & CAMPBELL, D. T. **Response Set in the F Scale** 129
CHAPMAN, L. J., & TAYLOR, JANET A. **Breadth of Deviate Concepts Used by Schizophrenics** 118
CHURCH, R. M. **Transmission of Learned Behavior Between Rats** 163
CLARK, J. W. See KAMIN, L. J.
COHEN, A. S. See STOTLAND, E.
CRASILNECK, H. B., & MICHAEL, CARMEN M. **Performance on the Bender Under Hypnotic Age Regression** 319
CROW, W. J., & HAMMOND, K. R. **The Generality of Accuracy and Response Sets in Interpersonal Perception** 384
CULBERTSON, FRANCES M. **Modification of an Emotionally Held Attitude Through Role Playing** 230
DAVIS, R. H., & HARRINGTON, R. W. **The Effect of Stimulus Class on the Problem-Solving Behavior of Schizophrenics and Normals** 126
DITTES, J. E. **Extinction During Psychotherapy or GSR Accompanying "Embarrassing" Statements** 187
EHRLICH, DANUTA, GUTTMAN, I., SCHONBACH, P., & MILLS, J. **Postdecision Exposure to Relevant Information** 98
ERIKSEN, C. W. See KUETHE, J. L.
ERRATA 428
FESHBACH, S., & SINGER, R. **The Effects of Personal and Shared Threats Upon Social Prejudice** 411
FINDLAY, D. C., & MCGUIRE, C. **Social Status and Abstract Behavior** 135
FINE, B. J. **Conclusion-Drawing, Communicator Creditability, and Anxiety as Factors in Opinion Change** 369
FREEMAN, J. T., & MAHER, H. **A Note on "Discrimination Reaction Time as a Function of Anxiety and Intelligence"** 134
FRENCH, J. R. P., JR., & ZAJONC, R. B. **An Experimental Study of Cross-Cultural Norm Conflict** 218
GLASSMAN, S. M. See HARDING, G. F.
GORDON, J. E. **Leading and Following Psychotherapeutic Techniques with Hypnotically Induced Repression and Hostility** 405

PAGE
GREENBANK, R. K. **The Wolf** ... 417
GRISELL, J. L. See ROSENBAUM, G.
GRUNES, WILLA F. **Looking at Occupations** ... 86
GUTTMAN, I. See EHRLICH, DANUTA.
GYNTHER, RUTH A. **The Effects of Anxiety and of Situational Stress on Communicative Efficiency** ... 274
HAMMOND, K. R. See CROW, W. J.
HARDING, G. F., GLASSMAN, S. M., & HELZ, W. C. **Maturation and the Spiral Aftereffect** ... 276
HARDY, K. R. **Determinants of Conformity and Attitude Change** ... 289
HARRINGTON, R. W. See DAVIS, R. H.
HEARN, G. **Leadership and the Spatial Factor in Small Groups** ... 269
HELZ, W. C. See HARDING, G. F.
HILDUM, D. C. See BROWN, R. W.
HOEHN, A. J. See SALTZ, E.
HOFF, D. B. See BRIM, O. G., JR.
HOVLAND, C. I., & PRITZKER, H. A. **Extent of Opinion Change as a Function of Amount of Change Advocated** ... 257
HOWES, D. **On the Relation Between the Probability of a Word as an Association and In General Linguistic Usage** ... 75
JACKSON, D. N. See MESSICK, S.; SOLLEY, C. M.
JACKSON, D. N., & MESSICK, S. J. **A Note on "Ethnocentrism" and Acquiescent Response Sets** ... 132
JACKSON, D. N., MESSICK, S. J., & SOLLEY, C. M. **How "Rigid" is the "Authoritarian?"** ... 137
JENSEN, A. R. **Authoritarian Attitudes and Personality Maladjustment** ... 303
JOHNSON, L. C. **Generality of Speed and Confidence of Judgment** ... 264
KAESS, W. A. See WITRYOL, S. L.
KAMENETZKY, J., & SCHMIDT, H. **Effects of Personal and Impersonal Refutation of Audience Counter-arguments on Attitude Change** ... 200
KAMIN, L. J., & CLARK, J. W. **The Taylor Scale and Reaction Time** ... 262
KELMAN, H. C., & PARLOFF, M. B. **Interrelations Among Three Criteria of Improvement in Group Therapy: Comfort, Effectiveness, and Self-Awareness** ... 281
KESSEN, W. **The Role of Experience in Judging Children's Photographs** ... 375
KLEINER, R. See PEPITONE, A.
KLEINMAN, M. L. **Psychogenic Deafness and Perceptual Defense** ... 335
KORCHIN, S. J. See BASOWITZ, H.
KORCHIN, S. J., & BASOWITZ, H. **Age Differences in Verbal Learning** ... 64
KORCHIN, S. J., & LEVINE, S. **Anxiety and Verbal Learning** ... 234
KRISTY, N. F. See CATTELL, R. B.
KUETHE, J. L., & ERIKSEN, C. W. **Personality, Anxiety, and Muscle Tension as Determinants of Response Stereotypy** ... 400
LEITER, R. A. See BROWN, R. W.
LEVINE, S. See KORCHIN, S. J.
LITTLE, K. B. See BRACKBILL, YVONNE.
MACKAVEY, W. R. See ROSENBAUM, G.
MCGAUGHRAN, L. S., & MORAN, L. J. **Differences Between Schizophrenic and Brain-Damaged Groups in Conceptual Aspects of Object Sorting** ... 44
MCGUIRE, C. See FINDLAY, D. C.
MAHER, B. A. **Personality, Problem Solving, and the Einstellung Effect** ... 70
MAHER, H. See FREEMAN, J. T.
MARTIN, B. See BERRY, J. L.
MASON, D. J. **Judgments of Leadership Based Upon Physiognomic Cues** ... 273
MATARAZZO, J. D. See SASLOW, G.
MATARAZZO, RUTH G. See SASLOW, G.
MAUSNER, B., & BLOCH, BARBARA L. **A Study of the Additivity of Variables Affecting Social Interaction** ... 250
MEELAND, T. See SEIDMAN, D.
MESSICK, S., & JACKSON, D. N. **Authoritarianism or Acquiescence in Bass's Data** ... 424
MESSICK, S. J. See JACKSON, D. N.
MESSICK, S. J. See SOLLEY, C. M.
MICHAEL, CARMEN M. See CRASILNECK, H. B.
MILLER, I. See SEIDMAN, D.
MILLS, J. See EHRLICH, DANUTA.
MORAN, L. J. See MCGAUGHRAN, L. S.
NUTHMANN, ANNE M. **Conditioning of a Response Class on a Personality Test** ... 19
PAPAJOHN, J. See SPIVAK, M.
PARLOFF, M. B. See KELMAN, H. C.
PEPITONE, A., & KLEINER, R. **The Effects of Threat and Frustration on Group Cohesiveness** ... 192

PAGE
PERKINS, J. C. See PODELL, L.
PHARES, E. J. **Expectancy Changes in Skill and Chance Situations** 339
PHILLIPS, JEANNE S. See SASLOW, G.
PICKREL, E. W. **Levels of Perceptual Organization and Cognition: Conflicting Evidence** 422
PODELL, L., & PERKINS, J. C. **A Guttman Scale for Sexual Experience—A Methodological Note** 420
PRITZKER, H. A. See HOVLAND, C. I.
RABINOWITZ, W. **Anality, Agression, and Acquiescence** 140
RAPHELSON, A. C. **The Relationships Among Imaginative, Direct Verbal, and Physiological Measures of Anxiety in an Achievement Situation** 13
RIOPELLE, A. J. See BEHAR, I.
ROSENBAUM, G., MACKAVEY, W. R., & GRISELL, J. L. **Effects of Biological and Social Motivation on Schizophrenic Reaction Time** 364
ROTHSCHILD, G. H. See SHAW, M. E.
SALTZ, E., & HOHEN, A. J. **A Test of the Taylor-Spence Theory of Anxiety** 114
SARASON, I. G. **Effect of Anxiety and Two Kinds of Motivating Instructions on Verbal Learning** 166
SASLOW, G., MATARAZZO, J. D., PHILLIPS, JEANNE S., & MATARAZZO, RUTH G. **Test-Retest Stability of Interaction Patterns During Interviews Conducted One Week Apart** 295
SCHAEFER, E. S., & BELL, R. Q. **Patterns of Attitudes Toward Child Rearing and the Family** 391
SCHMIDT, H. See KAMENETZKY, J.
SCHONBACH, P. See EHRLICH, DANUTA.
SCODEL, A., & AUSTRIN, H. **The Perception of Jewish Photographs by Non-Jews and Jews** 278
SEIDMAN, D., BENSEN, S. B., MILLER, I., & MEELAND, T. **Influence of a Partner on Tolerance for a Self-Administered Electric Shock** 210
SHAW, M. E., ROTHSCHILD, G. H., & STRICKLAND, J. F. **Decision Processes in Communication Nets** 323
SHOR, R. E. **Effect of Preinformation Upon Human Characteristics Attributed to Animated Geometric Figures** 124
SINGER, R. See FESHBACH, S.
SMITH, H. P. **The Effects of Intercultural Experience—A Follow-up Investigation** 266
SOLLEY, C. M., JACKSON, D. N., & MESSICK, S. J. **Guessing Behavior and Autism** 32
SOLLEY, C. M. See JACKSON, D. N.
SPENCE, D. P. **A New Look at Vigilance and Defense** 103
SPIVAK, M., & PAPAJOHN, J. **The Effect of the Schedule of Reinforcement on Operant Conditioning of a Verbal Response in the Autokinetic Situation** 213
STECKLER, G. A. **Authoritarian Ideology in Negro College Students** 396
STICE, G. F. See CATTELL, R. B.
STOTLAND, E., THORLEY, S., THOMAS, E., COHEN, A. S., & ZANDER, A. **The Effects of Group Expectations and Self-Esteem Upon Self-Evaluation** 55
STRICKLAND, J. F. See SHAW, M. E.
TAYLOR, JANET A. See CHAPMAN, L. J.
THOMAS, E. See STOTLAND, E.
THORLEY, S. See STOTLAND, E.
TROW, D. B. **Autonomy and Job Satisfaction in Task-Oriented Groups** 204
VEROFF, J. **Development and Validation of a Projective Measure of Power Motivation** 1
WEISS, W. **Opinion Congruence with a Negative Source on One Issue as a Factor Influencing Agreement on Another Issue** 180
WITRYOL, S. L., & KAESS, W. A. **Sex Differences in Social Memory Tasks** 343
ZAJONC, R. B. See FRENCH, J. R. P., JR.
ZANDER, A. See STOTLAND, E.

Announcing

PATTERNS OF CHILD REARING

by ROBERT R. SEARS
Stanford University

ELEANOR E. MACCOBY
Harvard University

HARRY LEVIN
Cornell University

A brilliant report on ways of bringing up children — based on data obtained from comprehensive interviews with a diversified sample of American mothers.

For the personal libraries of all who deal professionally with children.

For text use in courses in Personality, Parent-Child Relationships, Child Psychology, The Preschool Child.

1. Provides a fascinating report of how American mothers are rearing their children.
2. Describes and analyzes actual child-rearing practices obtained through personal interviews.
3. Contains numerous verbatim accounts by mothers colorfully describing their child-training experiences and problems.
4. Examines alternative training methods and shows their effects on the child.
5. Makes use of refined scientific techniques in obtaining information and evaluating data.
6. Contributes significantly to the theory of personality development.
7. Suggests promising avenues for further research in parent-child relationships.

No. pages: 560 No. illustrations: 43 Size: 6" x 9" Price: $5.25

ROW, PETERSON AND COMPANY

Evanston, Illinois

A new work from a
distinguished research team . . .

PHYSIQUE AND DELINQUENCY

By SHELDON and ELEANOR GLUECK, authors of "Delinquents in the Making," etc.

Drawing upon the mine of raw material provided by their comparative study of five hundred delinquent and five hundred non-delinquent youth in an earlier book, "Unraveling Juvenile Delinquency," the Gluecks have now correlated their psychiatric and social analyses with data on physique types. In this important new book, the authors point out that their latest findings indicate significant relationships between physique and behavior—that different body types respond differently to the forces of environment, giving clear indications of delinquency potential. They not only document this whole new area of research but suggest how the findings may be applied to the treatment and prevention of delinquency.

"Thanks to the Gluecks we now have more solid knowledge than ever before concerning the complex relationships that exist between physique and delinquency."—GORDON W. ALLPORT

$6.00

At your bookstore or from HARPER & BROTHERS, N. Y. 16